**Title**
How does learning shape us? Exploring mechanisms of plasticity in cognition

**Author**
Guerra-Carrillo, Belen Carolina

**Publication Date**
2018

# How Does Learning Shape Us?
## Exploring Mechanisms of Plasticity in Cognition

By
Belen Carolina Guerra-Carrillo

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy
in
Psychology
in the
Graduate Division
of the
University of California, Berkeley

Committee in charge:

Professor Silvia Bunge
Associate Professor Linda Wilbrecht
Assistant Professor Zachary Pardos

Spring 2018

# Abstract

# How Does Learning Shape Us? Exploring Mechanisms of Plasticity in Cognition

by

Belen Carolina Guerra-Carrillo

Doctor of Philosophy in Psychology

University of California, Berkeley

Professor Silvia A. Bunge, Chair

How do learning experiences shape our brain and cognitive skills? While there has been extensive evidence showing that learning experiences do have an impact on cognitive performance across the lifespan, we are still understanding how those changes take place. My dissertation work has sought to characterize the mechanisms that underlie improved performance resulting from different learning experiences. In Chapter 1, the general introduction, I present relevant background literature to provide a context for the rest of the chapters in my dissertation. In Chapter 2, I provide an overview of the type of changes to the functional connectivity of brain networks that result from sensory, motor, and cognitive learning experiences. The evidence from that body of work leads us to conclude that learning changes the way brain networks interact, such that BOLD fluctuations within the network engaged in the learning experience become more tightly synchronized at rest. Having established evidence for experience-dependent brain plasticity, we turn our attention to Chapter 3, where I examine the cognitive impact of broad learning experiences across five decades of life. I present findings suggesting that education has a greater influence on performance on assessments that measure more complex skills, like reasoning. Educational attainment also modulates ages of peak cognitive functioning, while it has little bearing on novel learning. Given that Chapter 2 and 3 provide evidence of the impact of learning on the brain and behavior, we then probe mechanisms of change. In Chapter 4, we present unified findings from a broad literature to explain how various eyetracking measures provide a relatively simple method to characterize cognitive mechanisms that support change as a result of maturation, which can be applied to study changes that occur as a result of learning. Using this eyetracking methodology, in Chapter 5 and 6, I provide empirical evidence about the mechanisms that support improvements in reasoning performance resulting from a real-world learning experience. We find that the young adults became more proficient at encoding, maintaining, and integrating visual relations after only a relatively short learning experience that emphasized reasoning about verbal relations and rules. Collectively, my dissertation work provides mechanistic insights about how learning experiences shape higher-level cognition.

# Table of Contents

# Acknowledgments

First, I would like to thank Silvia Bunge, my fantastic advisor, for her consistent encouragement and guidance, and for showing me how to turn my curiosity into testable questions. I would also like to thank the rest of my committee for their time, and useful discussions about this research: Linda Wilbrecht, your passion for science has been a source of inspiration over the years. Thank you for expanding my understanding of plasticity. Zachary Pardos, thank you for opening my eyes to a whole new approach to study learning and for showing me that research and application can go hand-in-hand.

I would also like to thank my wonderful colleagues who taught me so many things and who made science fun over the years: Maria Eckstein, Kiefer Katovich, Allyson Mackey, and Wei-Chun Wang. I would also like to thank Alison Miller Singley, my collaborator, grad school buddy, and kindred spirit in science. Thank you, Lisa Simpson and Keith Johnson, for all the fun science talks and beverage breaks. Lastly, thank you, Ariel Starr, my partner in crime at the lab, at conferences, and at the punchis. You have been an inspiration and the greatest friend.

I am also thankful to my family for their enormous support over the years. Daniel, my brother, and Nancy, my sister-in-law, for encouraging me to follow my dreams and sharing so many wonderful experiences with me in la playa and beyond. My mom, who is the sweetest and most supportive person I have had the pleasure of meeting. Tuca, the cuddliest of dogs, who kept me company in the long hours involved writing this work. Finally, Andrew, thank you for absolutely everything.

# Chapter 1

# General Introduction

Learning experiences keep our brain in flux throughout our lives. Over a century ago, scientists thought otherwise. The brain was predominantly viewed as being immutable to environmental influences, and the mind, as a set of inherited traits that could not be shaped through learning. A series of seminal experiments have since then helped shift our understanding of brain plasticity and the cognitive impact of learning experiences (for a review see: Greenough, Black, Wallace, Development, & Jun, 1987). Now, we have evidence that even playing video games (Bediou et al., 2018) or learning how to juggle (Draganski & May, 2008) can shape the structure and function of the brain. Indeed, we have learned a lot in the last 70 years. We know for example, that some systems do have narrow windows during development when they are highly sensitive to environmental input, but become relatively stable thereafter (Hensch, 2005). By contrasts, other systems retain a higher degree of malleability even in adulthood (Simone Kühn & Lindenberger, 2016; May, 2011). Greenough and colleagues (1987) proposed a compelling explanation that this latter capacity for change is an evolutionary solution that helps animals adapt existing systems to experiences that may occur at different points in their lives and that are unique to their environments. According to this view, changes that occur as a result of learning experiences that are unique to an individual (e.g., learning how to juggle), constitute a form of experience-dependent plasticity (Galván, 2010; Greenough et al., 1987). Over the years, we have continued to understand the nature of experience-dependent plasticity in various domains, including perceptual learning (Bavelier, Levi, Li, Dan, & Hensch, 2010; Bonaccorsi, Berardi, & Sale, 2014), motor skill acquisition (Galea, Vazquez, Pasricha, de Xivry, & Celnik, 2011; May, 2011), and changes in higher-order cognition. My dissertation work deals with the latter.

In this introduction, I will briefly review relevant background literature to provide a context for the chapters in my dissertation. I will first explain some key concepts and illustrate the relevance of this work. Then, I will provide examples of prior evidence for the malleability of cognitive functions, and discuss possible mechanisms of change. As I discuss these topics, I will highlight how my dissertation chapters expand on this prior work when appropriate, but I also provide a summary of the chapters towards the end.

## What is higher-level cognition and why does it matter?

Higher-level cognition is an umbrella term encompassing many functions, such as our ability to process information fluidly (processing speed), attend selectively to information relevant to our goals (selective attention), maintain and manipulate information in mind (working memory), integrate features of and consider relationships between stimuli (relational reasoning), among others. Each of these skills is considered a unique process, but all are highly interrelated (Mcardle, Ferrer-Caja, Hamagami, & Woodcock, 2002; Salthouse, 2005).

Relational reasoning is arguably the most complex of these skills, as it relies on our ability to attend to, process, and maintain information in mind (Kail, Lerv, & Hulme, 2016), but it also builds on other unique processes that allow us to identify patterns, encode relations, and make connections between otherwise unrelated information (Alexander, 2016; Halford, Wilson, & Phillips, 2010). These processes are referred to as relational thinking (Alexander, 2016), and it constitutes a foundational step in our ability to make analogies, draw inferences, and arrive at conclusions via deductions. Regardless of how complex or relatively simple one of these higher-cognitive skills may be, they all work together and constitute important aspects of human cognition.

These cognitive skills contribute to our everyday functioning, including our ability to learn in school. For example, a student would have a hard time following the lecture if they do not pay attention to the correct part of a slide being presented, or would likely make a mistake in their math homework if they fail to keep the correct numbers in mind involved in the computation at hand (Alloway & Alloway, 2010; Salthouse, 2005). Also, a student would grasp a new complex concept more easily if they can identify how that concept relates to something they have learned in the past (Vendetti, Matlen, Richland, & Bunge, 2015). In short, these skills help us discover and navigate our world more easily.

Given how important these skills are for our everyday functioning, you can imagine the promise it holds if we can understand whether and how they can be shaped by experiences. Thus far, we have gained a clearer understanding regarding the first part of this question. Indeed, there is ample evidence for the experience-dependent plasticity of higher-cognitive functions across the lifespan[1]. However, we are still working on understanding how these changes take place. That is the main focus of my dissertation work.

## Experience-dependent plasticity of higher-level cognition: evidence and approaches

We have evidence that cognitive functions are malleable and respond to a variety of learning experiences, ranging from prolonged multifaceted activities like schooling to short-term lab-based interventions that target one specific cognitive component, among others.

Examining the effects of natural learning experiences, such as the impact of formal schooling or majoring in a specific subject in school, have set a precedent for the idea that learning and complex environments shape cognitive functioning. The benefit of this approach is that it allows us to measure the effect of prolonged experiences that most people already naturally engage in. Some of the challenges have involved the correlational nature of the findings: the classic chicken-or-egg problem. There are a few ingenious instances where people have leveraged natural breaks in the schooling process to characterize the directionality of effects better. For example, in one study, Brod and colleagues (Brod, Bunge, & Shing, 2017) followed five years olds who were at the cutoff age to start first grade in Germany, some entered school

---

[1] I contributed to a book chapter (Blackwell, Rodriguez, & Guerra-Carrillo, 2014), where I discuss this work at greater length.

while other waited another year in kindergarten. At the end of the school year, the first graders had higher performance in a test of executive functions and increased engagement of brain regions that support the ability to sustain attention. These changes seemed attributable to the experience of schooling, rather than other changes in development or other individual differences. A recent meta-analysis of other quasi-experimental work have led to similar conclusions about the effect schooling on some reasoning measures (measured via I.Q tests), wherein changes in school reform prompted situations that permitted testing to a greater degree the directionality of effects (Ritchie & Tucker-Drob, 2017). These studies have provided compelling evidence for the cognitive effects of engaging in enriching experiences. The intensive, prolonged, and multifaceted nature of educational experiences provide a platform to also ask other exciting questions about the role of rich experiences in cognitive functioning. For example, which specific cognitive skills are most sensitive to educational experiences? What is the effect of these cumulative experiences in novel learning? And, what is the long-lasting impact of such prolonged experiences months, and even years, afterward? In Chapter 2, I discuss empirical work that sought to answer these questions.

Targeted interventions offer the advantage of allowing us to examine the effect of one specific learning experience in relation to other changes that may occur as a result of other complexities in the environment (Greenough et al., 1987; Tidwell, Dougherty, Chrabaszcz, Thomas, & Mendoza, 2014). This approach is not without its drawbacks. Lab-based interventions, for example, are difficult to implement, and there often ends up being a tradeoff between gaining selectivity and losing ecological validity. Many elegant arguments have been made for the importance of ecological validity (Moreau & Conway, 2014); one simple reason is that interventions can otherwise quickly become boring and irrelevant to a learner's goal. The learner needs to feel motivated and engaged in the program to even want to complete the study, and there is an argument to be made for the importance of novelty (Kolb et al., 2012).

A few of my favorite examples of targeted interventions that have overcome this challenge, while still retaining a respectable degree of selectivity, have involved playing fun games that emphasize a particular cognitive skill. The video-game work I alluded to earlier, for example, has shown that people who played several hours of first-person shooter games, which involve attending numerous moving targets, improved in various aspects of visuospatial attention compared to an active-control group who played life-simulation games (for a review see: Bediou et al., 2018). Improvements included better ability to sustain attention on target stimuli and quickly shift focus as needed to changing task demands on lab assessments. A separate body of work has shown that playing a variety of games that emphasize executive functions, for example, games that require conflict resolution and keeping bits of information in mind, have led to improved performance in assessments that measured similar skills (Goldin et al., 2014). The gains were even seen in academic performance for kids with low school attendance (Goldin et al., 2014). Games have also been an effective way to boost reasoning skills. Work from our own lab has shown that kids who played games that emphasized reasoning skills for eight weeks, such as those involving inferring rules, figuring out patterns, and solving visual logic puzzles improved on tests of relational reasoning (Mackey, Hill, Stone, & Bunge, 2011). By contrast, their classmates, who played games that emphasized making speeded decisions improved on measures of processing speed. Together, these experiments have provided important, ground-breaking work showing that short-term experiences that tax a particular skill lead to improved performance on unpracticed measures that emphasize the same skill.

These and several other studies have shown that cognition is malleable and sensitive to learning experiences. Some important questions remain, however. For one, we have tried to understand the factors that influence the magnitude of change in the desired skill (Diamond & Ling, 2016), and how improvements in the practiced context would translate to gain in other domains (Moreau & Conway, 2014; Noack, Lövdén, & Schmiedek, 2014). These two issues relate to the transfer of learning, which is the process by which previously acquired knowledge or skills may influence performance in a new situation (Woodworth & Thorndike, 1901). Evidence of far transfer has been sought after in a wide range of skill acquisition domains (Schunk, 1996), including in studies examining the plasticity of higher-order cognition. The extent to which gains observed from a cognitive intervention extend to improved performance on other measures (i.e., the degree of transfer), has been the subject of heated debate, at least for some domains (Melby-Lervåg & Hulme, 2013; Shipstead, Redick, & Engle, 2012). Conversely, there is a broader consensus that the degree of transfer would depend heavily on the extent to which the transfer and studied task rely on shared neural and cognitive mechanisms, and the extent to which these shared processes are engaged during the learning phase (Buschkuehl, Jaeggi, & Jonides, 2012; Lindenberger, Wenger, & Lövdén, 2017). However, we do not yet fully understand what are the putative mechanisms underlying change.

### Mechanisms of plasticity in higher-level cognition

Based on prior work, we know that there are factors that influence how much a person would benefit from a cognitive learning experience. Briefly, these include the age of a person (sorry, things start going downhill later in adulthood; Simone Kühn & Lindenberger, 2016), the length and practice intervals of the intervention (spacing practice helps; Wang, Zhou, & Shah, 2014), and, depending on the context, people's baseline levels of performance (the rich get richer, but not always; Diamond & Ling, 2016; Walberg & Tsai, 1983). Although we know these factors are associated with the magnitude of learning outcomes, we are still elucidating what are the mechanisms of change at the brain and cognitive level.

We have some evidence of experience-dependent brain plasticity at the level of gray and white-matter structure. For example, increases in gray matter volume in the hippocampus, cerebellum, and dorsolateral prefrontal cortex have been observed after two months of playing games that engage navigation and strategy (Kühn, Gleich, Lorenz, Lindenberger, & Gallinat, 2014). At the cellular level, there are several possible interpretations that could explain these changes, including synaptogenesis (Zatorre, Fields, & Johansen-Berg, 2012). However, that interpretation is likely an oversimplification, since the changes reported do not capture the time course of the effects of learning. There is evidence from the motor literature that at least three measurements are necessary to characterize the expected rise and fall in volume that occurs during skill acquisition in the human brain (Lindenberger et al., 2017). Even with three measurement timepoints, most studies could still miss any changes that occur at the level of gray matter (Lindenberger et al., 2017). Relatively few studies have characterized changes in white-matter structure following cognitive interventions. We have evidence that practicing reasoning with the Law School Admissions Test was associated with greater changes in white-matter microstructure of tracts in the frontoparietal network (Mackey, Whitaker, & Bunge, 2012). A plausible explanation is that the changes observed in the study correspond changes in myelination (Zatorre et al., 2012). However, due to methodological limitations of neuroimaging, we cannot conclusively pinpoint the cellular mechanisms of change, which could include changes in glial cells, vasculature, and water content in axons, to name a few (Mottershead et al., 2003; Zatorre et al., 2012). It would be exciting to see what more sensitive metrics of brain anatomy, including white matter imaging, reveal about the effects of learning experiences in the brain (Liu, Li, Tong, & Yeom, 2015; Sagi et al., 2012).

Other work has characterized experience-dependent brain plasticity with functional neuroimaging. For example, working memory training has led to both an increase and decrease in BOLD response in frontal and parietal regions typically engaged during performance of working memory tasks (Constantinidis & Klingberg, 2016). Based on findings with non-human primates that have shown that neural firing during retention periods increases with improved performance during training (Qi, Meyer, Stanford, & Constantinidis, 2011), we might speculate that changes in BOLD response relate to an enhanced representation of the stimuli. However, as Constantinidis & Klingber (2016) point out, there is not a straightforward relationship between measures of neuronal firing rates and BOLD respond, nor a clear way to interpret changes in BOLD activity at that level. Indeed, there has not yet been a unified explanation of what those changes may mean regarding brain plasticity, but we at least know that it is problematic to interpret them simply as signifying changes in neural efficiency (Poldrack, 2015). Aside from providing little insight regarding changes in the functional properties of the brain, we are also limited in how to interpret changes in fluctuations as they relate to changes in behavior. Considering that in the majority of the lifespan, there may be more limited structural changes that could be captured with the gross measures of brain structure currently available (Simone Kühn & Lindenberger, 2016), it holds promise to leverage co-fluctuations in brain activity between different regions to assess experience-dependent brain plasticity. This is a method my colleagues, and I have argued in the past, work that I discuss more fully in Chapter 1.

Another big unknown relates to the cognitive mechanisms that support improvements. As I have discussed earlier, higher-level cognition is multifaceted and relies on various processes. Behavioral measures only capture the output of all these computations, and thus, may not adequately help us characterize mechanisms of change. Furthermore, we are unable to draw inferences about changes in the cognitive processes from neuroimaging measures. One motivation to understand cognitive mechanisms is that they may be important contributors of plasticity in adulthood, a period during which it may be optimal to respond to new learning demands by repurposing established processes; for example, by engaging metacognitive strategies to supplement any shortcomings in performance (Lövdén, Bäckman, Lindenberger, Schaefer, & Schmiedek, 2010). In examining cognitive mechanisms, we can guide our hypotheses based on the relationship between different cognitive processes and how they are engaged in problem-solving. In Chapter 4, I present work my colleagues and I have unified to explain the utility of eyetracking to study mechanisms that support changes in cognition during maturation. Many of the same methods can be applied to examine cognitive plasticity that results from learning experiences. Indeed, in Chapters 5 and 6, I discuss how we have employed eyetracking metrics to examine cognitive mechanisms supporting changes in reasoning.

In sum, we know that people learn, but we are still figuring out how. Most of my dissertation work has sought to understand the mechanisms that support change.

## Overview of dissertation chapters

The next section, Chapter 2, provides an overview of changes in functional connectivity of large brain networks resulting from sensory, motor, and cognitive learning experiences. In that work, we argue for the utility of an imaging methodology to further our understanding of the plasticity of the networks that support higher-level cognition. Having established evidence for experience-dependent brain plasticity, we turn our attention to examining the effects of a prolonged and multifaceted learning experience on performance. In Chapter 3, I present empirical work characterizing the effects formal schooling on various domains of higher-level cognition. In that work, we had an unprecedented opportunity to examine the effects of education on performance and novel learning across five decades of the lifespan. Considering that Chapter 2 and 3 provide evidence of the impact of learning on the brain and performance, we turn our attention to probing mechanisms. In Chapter 4, we present unified findings from a broad literature to explain how various eyetracking measures constitute a relatively simple method for characterizing cognitive mechanisms of change. Using this methodology, in Chapter 5 and 6, I provide empirical evidence about the mechanisms that support improvements in reasoning performance in young adults who underwent a real-world learning experience.

# Chapter 2

## Resting-state fMRI: A window into human brain plasticity

Co-authors: Allyson P. Mackey [1], & Silvia A. Bunge [2, 3]

[1] Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology

[2] Department of Psychology, University of California, Berkeley

[3] Helen Wills Neuroscience Institute, University of California, Berkeley

**Permission to include co-authored material:** Granted

**Acknowledgments:** The authors thank Jesse Niebaum for helpful edits. We dedicate this review to the memory of William T. Greenough (1944-2013).

**Funding:** This work was supported by a graduate fellowship from the National Science Foundation to B.G., a National Research Service Award (NRSA) postdoctoral fellowship to A.M., and by a James S. McDonnell Scholar Award to S.B.

# Abstract

Although brain plasticity is greatest in the first few years of life, the brain continues to be shaped by experience throughout adulthood. Advances in functional magnetic resonance imaging (fMRI) have enabled us to examine plasticity of large-scale networks using blood-oxygen-level-dependent (BOLD) correlations measured at rest. Resting-state functional connectivity analysis makes it possible to measure task-independent changes in brain function, and therefore could provide unique insights into experience-dependent brain plasticity in humans. Here, we evaluate the hypothesis that resting-state functional connectivity reflects a repeated history of co-activation between brain regions. To this end, we review resting-state fMRI studies in the sensory, motor, and cognitive learning literature. This body of research provides evidence that the brain's resting-state functional architecture displays dynamic properties in young adulthood.

**Keywords**: brain plasticity, experience, resting-state fMRI, functional connectivity, training, practice, cognitive, sensory, motor, adult

# Introduction

Although the brain undergoes remarkable changes during early childhood, it retains the capacity to adapt to experience throughout life. Several decades ago, the late William T. Greenough proposed that brain plasticity is induced both by expected experiences shared among members of a species, or *experience-expectant* plasticity, and experiences that are specific to individuals, or *experience-dependent* plasticity (Greenough et al., 1987). Although these mechanisms likely fall along a continuum (Galván, 2010), the concept of experience-dependent brain plasticity provides the impetus for studying the brain changes that occur in adulthood as a result of the repeated engagement of specific neural systems through practice or training.

In the 65 years since Hebb first proposed the idea that patterns of coincident neuronal firing lead to structural changes that strengthen a synapse (Hebb, 1949), research in laboratory animals has led to remarkable progress in our understanding of brain plasticity at the cellular and systems levels (e.g., Blundon & Zakharenko, 2013; Hensch, 2005; Lisman, Schulman, & Cline, 2002). More recently, advances in structural and functional MRI data analysis have enabled us to measure experience-dependent brain plasticity in humans at the level of large-scale brain networks. In particular, Raichle's discovery of temporally correlated, low-frequency spontaneous fluctuations of BOLD signal across brain regions, known as *resting-state functional connectivity* (Raichle et al., 2001), provides an excellent opportunity to study brain plasticity in humans. Indeed, it has been hypothesized that these temporal correlations reflect the prior history of co-activation between brain regions (Buckner & Vincent, 2007; Dosenbach et al., 2007; Miall & Robertson, 2006). In this review, we evaluate the strength of the evidence for this claim.

Resting-state BOLD correlations are observed when subjects are instructed to relax inside the MRI scanner without engaging in a specific task. Temporal correlations do not appear to be random because patterns of connectivity have been reliably identified across studies and subjects (Damoiseaux et al., 2006; Smith et al., 2009). Further, patterns of correlation at rest follow along anatomical networks within primary sensory and motor cortices, as is the case of the somatomotor (SMN) and visual network (Figure 1). Resting state networks (RSNs) within association cortices include the dorsal attention, control/frontoparietal, salience, auditory, and default mode networks (DMN) (for a review see Buckner, Krienen, & Yeo, 2013). RSNs include but are not limited to areas that are monosynaptically connected. For example, right prefrontal cortex (PFC) and left parietal cortex show tightly correlated time courses at rest, but these regions are separated by at least two synapses.

The strength of correlations within and between networks has behavioral relevance. For example, visual connectivity is related to perceptual discrimination ability (Baldassarre, Lewis, Committeri, & Snyder, 2012), and frontoparietal connectivity is related to fluid intelligence and working memory (Cole, Yarkoni, Repovs, Anticevic, & Braver, 2012). These findings support the hypothesis that training-related changes in rs-FC *support* performance improvements; they are not only an epiphenomenon of repeated co-activation. The association to behavior also speaks to the relevance of examining changes in rs-FC in the context of plasticity and learning.

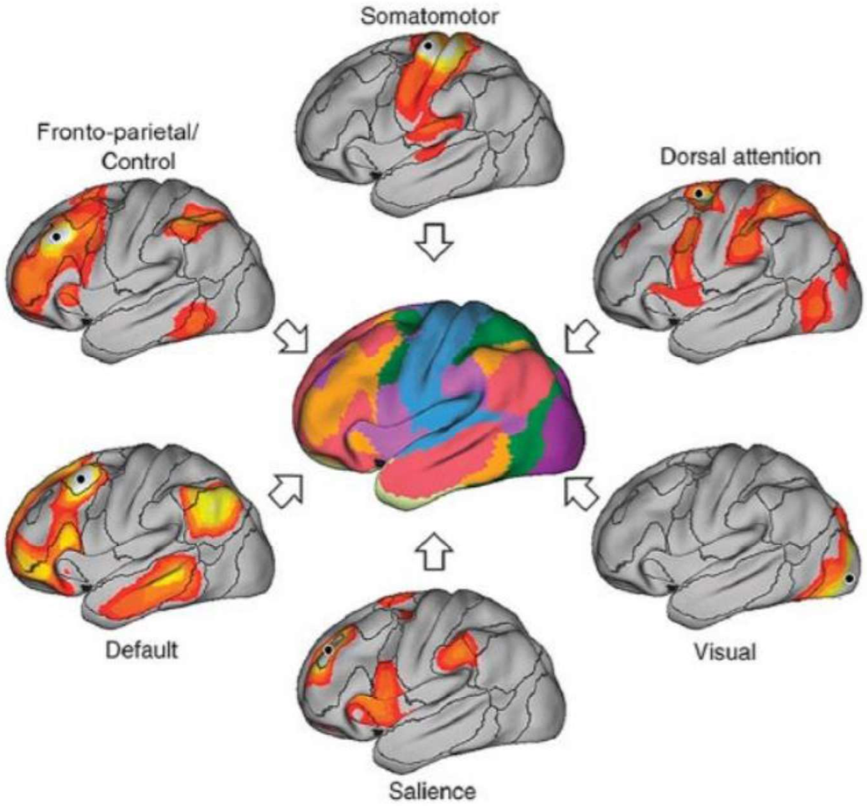


**Figure 1**. Resting state functional connectivity maps illustrates the organization of the human cortex into the major large-scale networks. The outer maps show, in red-yellow color scale, the regions that are functionality connected to the seed regions (dark circle). The visual and somatosensory networks show predominantly local connectivity. The default network, fronto-parietal/control network, and salience network present more distributed connectivity and lack strong coupling to sensory and motor areas. The map in the center displays a composite of the surrounding networks. We review findings from motor, sensory, and cognitive learning literature showing changes in the connectivity strength of these networks or between two independent regions. Adapted with permission from Buckner et al. (2013), *Nature Neuroscience* 16, 832-837.

This review begins with methodological considerations and then presents evidence supporting that rs-FC reflects experience-dependent plasticity by summarizing findings from rs-fMRI studies involving healthy young adults. It will also present studies that have used neuroscientific methods to induce plasticity, such as transcranial direct current stimulation (tDCS). This review will not cover the effect of medication on resting-state BOLD correlations nor review findings from developmental samples or clinical populations (for a recent review covering studies on brain injury, (see Gillebert & Mantini, 2013).

## Methodological Considerations

In order to evaluate the evidence in support of dynamic changes in rs-FC, we will first briefly address important methodological considerations that could bias the signal and interpretation of rs-FC. We will also consider physiological factors that impact connectivity patterns measured at rest.

**Experimental Design.** The instruction a participant receives prior to scanning could influence resting-state connectivity patterns. For example, Benjamin and others (2010) compared the connectivity in the DMN after subjects received instructions either to attend to the scanner background noise, not to attend to that noise, or simply relax and remain still. When subjects were asked to attend or ignore the noise compared with asked to relax, there was greater activation in dorsomedial PFC. The studies included in this review have taken precautions with respect to the instructions provided to participants. In some studies, however, participants received instructions to either keep their eyes open, closed, or fixated in a simple visual stimulus (e.g., a crosshair). Even though these factors modulate the strength of the connectivity of resting state networks (Patriat et al., 2013), researchers minimized these confounds by giving participants the same instructions in the baseline and post-training scanning session.

In addition to the instructions a participant receives for the rs-fMRI scan itself, activities performed immediately preceding the resting scan run could also impact the rs-FC signal. For instance, slow fluctuations occurring during cognitively demanding tasks have been found to have a greater delayed recovery period, indicating that *traces* of BOLD response to the cognitively demanding tasks do not subside, even after task completion (Barnes, Bullmore, & Suckling, 2009). Tung and others (2013) showed that even a task as simple as pressing a button for a fixed duration could significantly affect rs-FC correlations and fluctuations between regions and these changes return to baseline only after several minutes. The slow recovery period from a task could impact BOLD correlations observed during resting scans that are acquired directly after the task. Although the influence of prior cognitive tasks could bias rs-FC, there are pragmatic reasons that an experimenter would choose to acquire resting scans at the end of an experimental session or in addition to other tasks, such as the considerable cost for running a longitudinal imaging study. Of the studies included in this review, researchers conducting rs-FC at the end of the sessions minimized potential confounding effects by keeping the sequence of the tasks fixed across participants and experimental conditions (e.g., Mackey, Miller Singley, & Bunge, 2013; Powers, Hevey, & Wallace, 2012; Urner, Schwarzkopf, Friston, & Rees, 2013). Nonetheless, we encourage researchers to attempt to control as much as possible the level of cognitive demand required by participants before or during the scanning session.

Scan duration could also influence the reliability of rs-FC. Birn and others (2013) investigated the rs-FC in time series lengths varying from three to 27 minutes in three-minute increments. The length of the scan significantly increased the consistency of the rs-FC measured within a session, as measured by an increase of the interclass correlation (ICC); this increase slowed down after nine minutes and plateaued at 13 minutes. However, Van Dijk and others (2010) acquired scans varying in length between two to 12 minutes and found that approximately five minutes of data were sufficient to obtain moderate to high reliability.

**Data Analysis.** In addition to design confounds, data processing steps could also influence the strength or directionality of the correlations between regions. For example, the use of global signal removal has been subject to debate because this preprocessing step could bias the rs-FC signal by introducing anticorrelations. These negative correlations arise due to the mathematical properties of global signal regression, which leads to a negative mean correlation value in seeded connectivity analyses (Murphy, Birn, Handwerker, Jones, & Bandettini, 2009). Some researchers argue that global signal regression is a useful step because it removes high correlations driven by physiological noise. These researchers have opted to perform separate correlation analyses using only positive, only negative, and only absolute values (Cole et al., 2012). Others have proposed that anatomically and temporally constrained methods of physiological noise reduction make it possible to observe real, not artifactual, anticorrelations between networks (Chai, Ofen, Gabrieli, & Whitfield-Gabrieli, 2014). These anticorrelations are believed to reflect behaviorally relevant network segregation (Behzadi, Restom, Liau, & Liu, 2007).

**Physiological Confounds.** In addition to factors related to experimental design and data analysis, the physiological state of the participants can also influence the connectivity patterns observed (for review, see (Duncan & Northoff, 2013). Stress and sleep patterns in the days leading up to the resting state scan could affect the architecture of the RSNs and the strength of rs-FC. Such effects have not been widely reported, but these factors are of prime concern given their well-documented influence in plasticity (McEwen & Morrison, 2013; Walker & Stickgold, 2006). In a recent study (Vaisvaser et al., 2013), a non-clinical sample of young males was subjected to a Trier Social Stress Task, and rs-FC changes were analyzed from a scan acquired before the stressor, one scan recorded shortly after the stress induction, and one acquired after a 90-minute intermission. Although by the third scan most transient changes in the rs-FC were not significantly different from baseline, the strength of the correlation between the posterior cingulate cortex and hippocampal regions of interest (ROI) remained significantly stronger. The change in rs-FC connectivity between these regions was also correlated with changes in subjective perception of stress.

The effects of stress have also been observed over more prolonged periods of exposure to stressors. For example, Soares and others (2013) measured the effects of stress on DMN functional connectivity in medical interns over a three-month preparation period for residency examination. Compared with the changes observed in medical interns that were not preparing for the exam, the stressed group showed stronger connectivity within RSNs, such as the somatomotor network and DMN. These participants also showed increased cortisol response and self-reported feelings of stress after this period of preparation. In this study, the effects of studying for the exam remain elusive because they were not reported or correlated with the functional data. Although the effects of chronic and acute stress produce different physiological effects, it would be of interest to use a similar analytic approach to compare the results of the Soares and Vaivaser groups and to gain a better understanding of the different effects of acute and chronic stress on rs-FC.

Partial sleep deprivation has also been reported to affect the properties of rs-FC. Following a night of no sleep, young adults with otherwise normal sleep routines showed a decreased coupling within the DMN compared with controls whose sleep was undisturbed (De Havas, Parimal, Soon, & Chee, 2012). Similar results were observed in a study that measured FC from resting periods of a memory task in participants who experienced a night of sleep deprivation (Gujar, Yoo, Hu, & Walker, 2009). The connectivity patterns showed significantly less deactivation of the dorsal anterior cingulate cortex (dACC) with the DMN.

**Summary.** The studies reviewed above indicate that experiment design decisions that could impact the cognitive demand required by subjects, subject-level physiological characteristics, and data analysis steps are important factors to consider in evaluating rs-fMRI studies of brain plasticity, as they could confound the effects of training at the individual and group levels. At the end of this review, we provide recommendations for future investigations.

## Experience-dependent Changes in rs-FC

We begin by presenting findings from sensory and motor training studies, given their prominence in the study of plasticity. We will then review findings from the cognitive training literature, which has been traditionally more focused on human studies. In studies from these three domains, we expect to see changes in rs-FC between regions and within networks that are implicated in the trained domain. A summary of the results from the studies covered in this section is presented in Table 1.

**Plasticity in the Sensory Domain.** Sensory plasticity can take place relatively fast, and rs-FC could be a valuable index of the functional changes taking place between regions that support the sensory processing. This idea is supported by research conducted by Powers and others (2012), who trained participants in multisensory temporal integration by providing accuracy feedback to the participant's judgment of whether an auditory stimulus was played synchronously with a visual stimulus that was displayed 150 ms before or after the onset of the sound in 50 ms intervals. Performance on this task improved significantly after one training session, and the rs-fMRI scans revealed increased coupling between the posterior superior temporal sulcus (pSTS), secondary auditory cortex, superior colliculus, and superior cerebellum, among other regions (Figure 2). These areas support the audiovisual perception and integration capacity, as well as the timing of the integration.

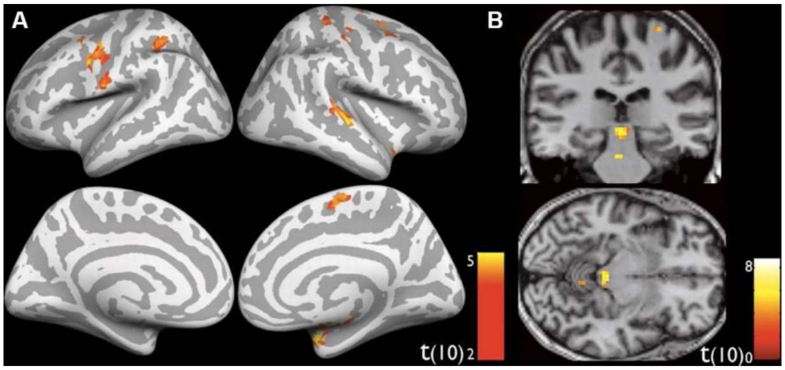


**Figure 2**. Data from study 3 in Table 1. Resting-state functional connectivity changes after audiovisual integration training. (A) The functional map illustrates the cortical areas with greater connectivity increase with the posterior superior temporal sulcus (pSTS). (B) This map shows a connectivity increase between the pSTS and superior colliculus. These regions support sensory integration and the connectivity increases illustrate that changes observed at rest are seen in regions that support the trained function. Adapted with permission from Powers et al. (2012). *Journal of Neuroscience*, May 2, 2012 • 32(18):6263– 6274.

Sensory plasticity has also been observed during motion perception task training. In this study, Urner and others (2013) investigated the lasting effects of visual learning by acquiring resting scans immediately after the training session and 24 hours later and used stochastic dynamic causal modeling to test for changes in the connectivity at these two time points. The day after training, the best-fitting model showed a lasting increase in the bidirectional connectivity between the hippocampus and striatum, which was interpreted as an index of consolidation. These findings suggest that even after a training period is over, changes in rs-FC can be observed. It is unknown whether these functional connections remain present for longer periods after learning.

The plasticity of other sensory modalities, mainly somatosensory input, has been examined through the use of acupuncture (Dhond, Yeh, Park, & Kettner, 2008) and electrical nerve stimulation (Klingner, Hasler, Brodoehl, Axer, & Witte, 2012). These studies have shown changes in rs-FC after a single session, either in canonical RSNs or the connectivity between regions of the secondary somatosensory cortex, thalamus, and association areas.

These studies provide evidence that repeated training can induce plasticity in the specific networks and brain regions involved in sensory processing. The regions that exhibited change in rs-FC were consistent with the sensory experience manipulated in the training paradigm. Interestingly, these changes could be observed even after a single training session, which may suggest rapid plasticity of the regions supporting sensory processing; however, it would be important to examine whether the changes resulting from repeated experience are sustained from longer periods of time or are only transient adaptations that result from the short length of training. A distinction between these two effects was tested by Urner and others (2013) but only after a 24-hour delay.

**Plasticity in the Motor Domain.** The use of rs-FC has provided opportunities to study the effects of motor learning in areas that are not classically defined as *motor regions* (i.e., support limb movement) but contribute to developing a complex motor skill. For example, changes in the coupling between motor regions and visual areas were observed after participants learned to reach for a target guiding their movements with indirect visual feedback and readjusting against resistance (Vahdat, Darainy, Milner, & Ostry, 2011). Vahdat and colleagues were able to dissociate rs-FC changes related to training that were mutual or exclusive to perceptual and motor learning. As shown in Figure 3, the changes in rs-FC between primary motor cortex and cerebellum were exclusive to the motor index of learning. These findings illustrate the benefit of using rs-FC measures to dissociate changes that occur in different systems that are not contingent on a specific task.

In addition, rs-FC can also reflect the progression of the neuroplastic changes that occur throughout a training program. Ma and others (2012), for example, trained participants to repeat a finger tapping sequence with their non-dominant hand. The strength of rs-FC between the right primary somatosensory cortex and right supramarginal gyrus increased after two weeks of training and decreased from the second to the fourth week. Interestingly, behavioral learning was only observed after the second week and was stabilized by the fourth week. A similar decrease in connectivity was reported in a separate study in which participants learned to manipulate a tool with their non-dominant hand (Yoo, Sohn, & Jeong, 2013). After eight weeks of practice, the correlations decreased within the sensory-motor network components, specifically between left primary motor cortex, supplementary motor area, and primary sensory cortex. Similarly, there was a decrease in rs-FC between regions that were initially correlated with manipulation of the tool, such as between right supramarginal gyrus and right premotor cortex. The authors suggested that this change in correlations reflects an enhanced efficiency in the functional network supporting motor control. However, the change in the directionality of rs-FC is not always consistent across motor learning paradigms. For example, rs-FC from training involving learning to balance an unstable structure showed increased coupling between bilateral supplementary/pre-supplementary motor areas and right ventral premotor cortex from baseline to the third week of training (Taubert, Lohmann, Margulies, Villringer, & Ragert, 2011).

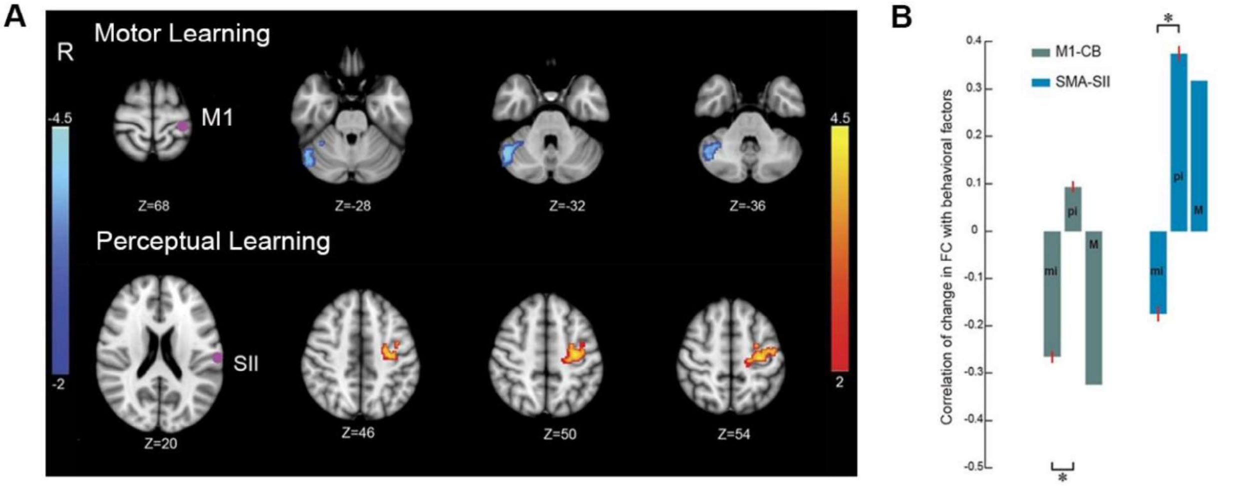


**Figure 3**. Data from study 10 in Table 1. Resting-state functional connectivity changes after motor reaching specific to perceptual and motor improvements. (A) Each row represents changes in connectivity in relation to motor and perceptual learning. Left column show the location of the seed ROIs; other columns show clusters with significant change in connectivity with the seed. Z-score maps show increase in positive correlations (shown in red-yellow) and negative correlations (shown in dark-light blue) from training day (day 1) to post-learning testing (day 2) with respect to learning. (B) Mean correlation between individual changes in FC are shown in the bar graph, from left to right, for sensory (pi), motor (mi), and mutual (M) indices of learning, respectively. Green bars indicate the connectivity between primary motor cortex (M1) and Cerebellum (CB). Blue bars illustrate the connectivity between supplementary motor area (SMA) and secondary somatosensory cortex (SII). Adapted with permission from Vahdat et al. (2011) *Journal of Neuroscience*, November 23, 2011 • 31(47):16907–16915 • 16907

Resting state FC from motor training studies indicates that the changes associated with the development of a motor skill are not limited to the regions that support the movement initiation but rather share intrinsic correlations with regions implicated in the perceptual components of the task. It is important to note that in these studies, the networks that exhibited changes were those expected to be engaged when performing the motor skill (Vahdat et al., 2011).

**Plasticity in the Cognitive Domain.** Behavioral and neuroimaging research examining the plasticity of higher cognition has aimed to develop and implement training regimes that could ameliorate or improve performance in a particular cognitive process, such as attention (Bavelier, Green, Pouget, & Schrater, 2012; Neville et al., 2013), reasoning (Bergman Nutley et al., 2011; Mackey et al., 2011, 2012), and working memory (Dahlin, Neely, Larsson, Bäckman, & Nyberg, 2008; Klingberg et al., 2005; Söderqvist et al., 2012). As the field has progressed, emphasis has also been placed in understanding the mechanism underlying plasticity of the systems supporting these functions (Mcnab, Nix, Hauth, Jorgensen, & Bastiani, 2009). However, many findings are still subject to debate, such as individual differences determining who would benefit from an intervention showing transfer of learning to untrained tasks (Melby-Lervåg & Hulme, 2013; Shipstead et al., 2012). We argue that the analysis of rs-FC data could illuminate answers to these debates if we consider this connectivity as reflecting how networks are reconfigured as a result of repeated practice. The studies presented below demonstrate an important first step supporting the possibility that rs-FC measures could shed light on the effects of cognitive training, since they show that repeated engagement of a neural circuit supporting higher cognition is reflected in changes in rs-FC.

To investigate changes in rs-FC after repeated experience with reasoning problems, Mackey and others (2013) recruited two groups of age- and IQ-matched pre-law students, in which only one group received preparation for the law school admission test (LSAT). The preparation course consisted of 100 hours of training over a three-month period, wherein two-thirds of the curriculum was devoted to solving reasoning problems. After the three months of training, the experimental group showed an increase in fronto-parietal and parietal-striatal rs-FC within and between hemispheres, as shown in Figure 4. Notably, these changes were linked to larger LSAT improvements. Interestingly, increased connectivity between the striatum and PFC ROIs was associated with smaller behavioral improvements, which suggest that differences in the changes in rs-FC could also serve to reflect an individuals' success in the training program.

Changes in rs-FC were also observed after four weeks of working memory training (Takeuchi et al., 2013). Participants were asked to perform a series of tasks that required them to maintain, update, or reorder verbal or visuospatial information. The group of participants that received training showed decreased coupling between mPFC and precuneus. Changes in the connectivity pattern were not correlated with behavioral performance on the working memory measures or reasoning tasks, which were considered transfer measures. The lack of correlation between rs-FC and behavior may be explained by the low between-subject variability in improvements after training. In a separate study by the same group (Takeuchi et al., 2013), participants were trained to multitask by performing exercises in which they had to process simultaneously and respond to different target stimuli, such as spoken numbers played to the left ear and vowels played to the other. The results showed decreased rs-FC between dorsolateral PFC and ventral ACC, regions involved in cognitive flexibility, but these changes were not correlated with behavioral improvements in multitasking and related cognitive process.

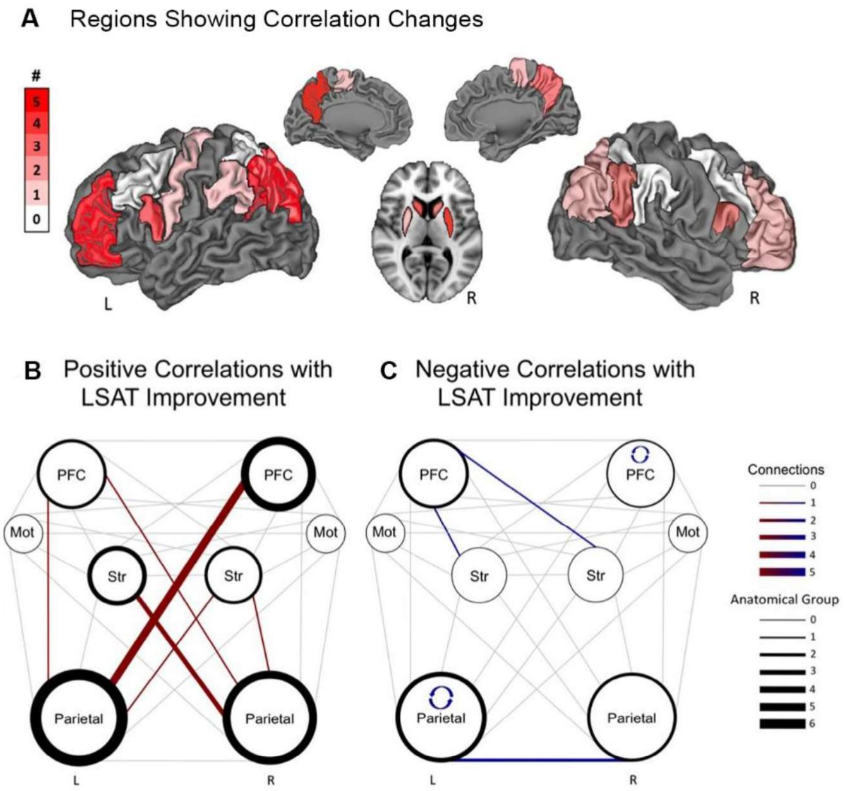


**Figure 4**. Data from study 12 in Table 1. Resting-state functional connectivity (FC) changes and correlation with behavioral improvements after preparation for the Law School Admission Test (LSAT), which strongly emphasizes reasoning. (A) Number of increased pairwise correlations for 11 brain regions most often recruited during reasoning performance (Prado et al., 2011), including rostrolateral prefrontal cortex and angular gyrus. (B-C) These diagrams indicate the increase in pairwise FC that were positively (B) and negatively (C) correlated with LSAT improvements. The thickness of the lines represents the number of connections with significant correlations. For more details, see original study. Adapted with permission from Mackey et al. (2013), *Journal of Neuroscience*, March 13, 2013 • 33(11):4796 – 4803

In these cognitive training studies, changes in rs-FC were studied in the context of experience-dependent plasticity and were also related to behavioral improvements in the trained domain (although with some mixed success, such as Takeuchi et al., 2013). Small sample sizes may partially explain the lack of behavioral associations, and further studies should be conducted to further confirm correlations between changes in rs-FC and behavior. The relationship between the distinct patterns of rs-FC and behavioral improvements reported by Mackey and others (2013) indicates the potential of rs-FC as a measure of learning success. It remains to be tested, however, whether BOLD correlations at rest could also be used to predict performance in an untrained cognitive domain and whether these correlations can serve as an index of individual differences in the benefits obtained from training.

**Table 1**. Summary of results from studies reporting plasticity of resting-state connectivity after motor, sensory, and cognitive training.

| Domain | Publication | Training | Duration | Control Group | $N$ | Analysis | Connectivity Changes | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | Increases | Decreases | No change |
| Sensory | 1) Dhond, 2008 | Acupuncture | 1 day; 150s | N | 15 | ICA | multiple DMN & SMN | | |
| | 2) Klingner, 2012 | Electrical nerve stimulation | 1 day; 16m | N | 12 | SCA | somatosensory cortices-(many: motor, visual) | somatosensory cortices-thalamus | |
| | 3) Powers, 2012 | Audiovisual simultaneity | 1 day; 1h | N | 11 | SCA | posterior superior temporal sulcus-auditory cortices | | |
| | 4) Urner, 2013 | Motion judgment | 1 day; 90m | N | 11 | ROIs | hippocampal-striatal | | |
| | 5) Ventura-Campos, 2013 | Phoneme discrimination | 2 weeks; 6h | N | 19 | SCA; ICA | insula-Salience network | insula-FPN | multiple FPN; temporal seed |
| Motor | 6) Albert, 2009 | Visuomotor adaptation | 1 day; 11m | A | 24 | ICA | FPN; cerebellar IC | | |
| | 7) Sami, 2013 | Visuomotor adaptation; pressing sequence | 1 day; 1.5h | A | 24 | GTA | strength (global, visual, frontal); degree (global, frontal, cerebellum) | path length (global, frontal, cerebellum, basal); centrality (frontal, parietal) | global efficiency & small worldness; local modularity |
| | 8) Ma, 2012 | Tapping sequence | 4 weeks; 7h | A | 10 | ICA | R parietal clusters | R parietal clusters | |
| | 9) Taubert, 2011 | Balancing | 6 weeks; 4.5h | NC | 28 | SCA; GTA | multiple SMN; centrality of motor cortices | | |
| | 10) Vahdat, 2011 | Field-force reach | 2 days; 2h | N | 13 | ROIs | multiple SMN | cerebellar-fronto motor; cerebellar-parietal | DMN |
| | 11) Yoo, 2013 | Tool manipulation | 1 month; 30h | N | 8 | ROIs; ICA | | SMN; intra parietal; parietal-cerebellar | DMN |
| Cognitive | 12) Mackey, 2013 | Reasoning | 3 months; 100h | NC | 48 | ROIs | multiple FPN; parietal-striatal | | |
| | 13) Takeuchi, 2012 | Working Memory | 27 days; 9-27h | NC; A | 81 | SCA | | medial PFC-precuneus | |
| | 14) Takeuchi, 2013 | Multitasking | 27 days; 9-27h | NC; A | 81 | SCA | | R dorsolateral PFC-vACC | |
| | 15) Xue, 2011 | Integrative body-mind | 1 month; 15h | A | 32 | GTA | nodal efficiency & degree value in L ACC | nodal efficiency & degree value in fronto motor areas | nodal efficiency & degree value in R ACC |

Note. Control Group: A= active, NC=no-contact, N=none. Analysis: ROIs= ROI correlations, SCA=Seed correlations, ICA=Independent component, GTA=graph theory. DMN=Default Mode Network, Fronto-Parietal Network (FPN), Somato-Motor Network (SMN), R=right, L=left, PFC= prefrontal cortex, ACC= anterior cingulate cortex. Brain regions are organized into resting state networks when possible. All connectivity increases and decreases are statistically significant at least at $p < .05$

## Discussion

The studies presented in this review support the hypothesis that the dynamic nature of resting-state functional connectivity reflects experience-dependent plasticity and illustrate that rs-FC patterns change after repeated experience with sensory, motor, or cognitive tasks. Similar changes in resting state connectivity have also been reported from cross-sectional studies comparing the effects of expertise, such as with experienced meditators (Taylor et al., 2013), trained musicians (Luo et al., 2012), and expert athletes (Di et al., 2012).

These findings raise questions about the nature of resting state architecture and the plasticity of neural systems. One important question is whether the changes observed in training studies in each domain are confined to the specific resting state networks associated with training demands. In the motor training literature, changes were confined to regions believed to support motor functions but were not exclusive to the canonical motor regions. Additionally, changes were observed in other networks that were also involved in the training (e.g., sensory cortices). However, testing for double dissociations of training effects is not widely performed. Vahdat and others (2011) tested the changes in connectivity exclusive to motor or perceptual learning. Furthermore, they tested for the changes observed at rest to be selectively present in the somatomotor network by examining plasticity in networks that were hypothesized to be unrelated to their training task (e.g., DMN). As expected, only the strength of the connectivity in the somatomotor network changed as a result of motor practice. Similar results were observed by Yoo and others (2013), who also observed changes specific to the motor network but not the DMN. It would be of interest for future work to test for these double dissociations more regularly, as these tests provide stronger casual evidence that the changes observed at rest reflect experience-dependent plasticity in specific circuits.

A second empirical question that could be addressed with rs-FC is whether distinct networks have comparable potential for malleability. Complementary use of neuroscientific tools, such as with tDCS, with rs-FC will allow these comparisons. The use of this manipulation would help researchers overcome the challenge of developing a training program that would equally target different networks. For example, Polanía and others (2011) found that after 10 minutes of bipolar tDCS with anodal stimulation to the left primary motor cortex (M1) and cathode stimulation to the contralateral frontopolar cortex, there was a strengthening of the intrinsic coupling between motor regions (left premotor, lM1, and left parietal cortex), and this change was accompanied by an increase in the efficiency of connectivity between the region that received the anodal stimulation and the rest of the brain. Such stimulation techniques could be used not only to compare the malleability potential of different networks but also the timing of these effects.

The studies reviewed from each modality (sensory, motor, and cognitive) indirectly address the belief that the rate of plasticity varies depending on which modality is being trained. This hypothesis can be inferred from the training duration necessary to induce changes in rs-FC. For example, studies examining plasticity of higher cognition showed changes after nearly 100 hours of training (e.g., Mackey et al., 2013), whereas those in the sensory domain identified plastic alterations after only one hour of training (e.g., Powers et al., 2012). To test this hypothesis directly, multiple scanning sessions at different intervals would be necessary to capture the changes occurring as a result of training.

An important question to consider is how changes observed in rs-FC are related to changes in behavioral performance. The results from the studies included in this review indicate that changes in rs-FC were strongly correlated with changes in behavior in the trained domain (e.g., Taubert et al., 2011; Mackey et al., 2013). Interestingly, in the studies that acquired more resting state data during the course of the study, certain patterns of change in rs-FC were detected prior to the actual behavioral improvement (Ma, Narayana, Robin, Fox, & Xiong, 2012; Taubert et al., 2011). Resting state FC could reflect the plasticity that can be observed prior to behavioral changes, although a link between behavioral improvements and changes in rs-FC was not observed in all studies (e.g., Takeuchi et al., 2013). In addition, correlations observed at rest have also served to predict behavioral gains from training paradigms, similar to how structural and task-based functional measures have been used in the past (for a recent review, see (Zatorre et al., 2012). For example, prior to training, Ventura-Campos and others (2013) were able to predict which participants would exhibit greater learning success after a training requiring the identification of non-native phonemes. The participants exhibiting greater learning showed stronger rs-FC between language-specific regions at baseline compared to poorer learners. The correlations of rs-FC could serve as an additional indicator of future behavioral gains and elucidate individual differences observed in training studies.

As mentioned in the cognitive plasticity section, rs-FC could provide valuable insights to test for transfer. One can speculate that widespread changes to a network are more likely to support transfer on untrained tasks than more localized changes. Individual difference analyses could provide insights into the mechanisms that underlie transfer, and predict which transfer tasks are likely to show behavioral improvements.

Functional connectivity observed at rest could be used to answer the empirical questions discuss thus far. It is important, however, for future research to also address the mechanisms underlying the changes observed in these intrinsic fluctuations. Computational and comparative work is needed to formulate possible cellular underpinnings of the changes observed in rs-FC, similar to the work that has been done to understand the cellular basis of structural plasticity in humans (for a review, see Zatorre et al, 2012). These investigations could also facilitate the understanding of what increases or decreases in rs-FC after a period of training specifically indicate.

In addition to elucidating the mechanisms underlying the changes in rs-FC, it is recommended that future work follows standards to control for the possible confounding variables discussed previously. Importantly, given the documented effects of sleep and stress in plasticity and rs-FC (De Havas et al., 2012; Soares et al., 2013; Vaisvaser et al., 2013), these physiological variables should be reported more consistently in the literature to control for potential individual and group differences. Although measuring physiological states is cost-intensive when these variables are considered of no-interest, it would be valuable to at least collect self-report measures. It would also be beneficial to monitor and report as much as possible other methodological considerations that could confound within- and between-subject comparisons, such as motion during scanning and cognitive demands prior to the scanning session. Considering these factors could facilitate comparisons between studies, although when comparing data obtained from different training paradigms, the analysis approach would also need to be equated. These comparisons could help advance the understanding of the efficacy of different training programs and rs-FC plasticity.

The findings reviewed here support the concept that rs-fMRI serves as an effective measure of plasticity; rs-fMRI appears to reflect a record of repeated synchronized activation between regions, as seen from correlation changes occurring in regions closely related to the trained task. The standardization of methodological approaches and better knowledge of the cellular underpinning of the changes observed in rs-FC may allow comparisons between studies and facilitate a better understanding of the impact of training on the connectivity observed at rest and provide greater insights into experience-dependent plasticity.

**Next chapter.** Having established evidence for experience-dependent brain plasticity, we turn our attention to examining the effects of a prolonged and multifaceted learning experience on performance. In Chapter 3, I present empirical work characterizing the effects formal schooling on various domains of higher-level cognition. In that work, we had an unprecedented opportunity to examine the effects of education on performance and novel learning across five decades of the lifespan.

# Chapter 3

## Does higher education hone cognitive functioning and learning efficacy? Findings from a large and diverse sample

Co-authors: Kiefer Katovich[1], & Silvia A. Bunge[2, 3]

[1]General Assembly Space, Inc., San Francisco, CA, United States of America

[2]Department of Psychology, University of California at Berkeley, Berkeley, CA, United States of America

[3]Helen Wills Neuroscience Institute, University of California at Berkeley, Berkeley, CA, United States of America

**Permission to include co-authored material:** Granted

**Acknowledgements:** We thank Ariel Starr and Yana Fandakova for their feedback, and Lumos Labs for providing de-identified data and allowing K.K. to assist with analysis.

**Funding:** National Science Foundation Graduate Research Fellowship (B.G.C.), the James S. McDonnell Foundation Scholar Award (S.A.B.), and the Jacobs Foundation Advanced Career Research Fellowship (S.A.B.).

# Abstract

Attending school is a multifaceted experience. Students are not only exposed to new knowledge but are also immersed in a structured environment in which they need to respond flexibly in accordance with changing task goals, keep relevant information in mind, and constantly tackle novel problems. To quantify the cumulative effect of this experience, we examined retrospectively and prospectively, the relationships between educational attainment and both cognitive performance and learning. We analyzed data from 196,388 subscribers to an online cognitive training program. These subscribers, ages 15-60, had completed eight behavioral assessments of executive functioning and reasoning at least once. Controlling for multiple demographic and engagement variables, we found that higher levels of education predicted better performance across the full age range, and modulated performance in some cognitive domains more than others (e.g., reasoning vs. processing speed). Differences were moderate for Bachelor's degree vs. High School ($d = 0.51$), and large between Ph.D. vs. Some High School ($d = 0.80$). Further, the ages of peak cognitive performance for each educational category closely followed the typical range of ages at graduation. This result is consistent with a cumulative effect of recent educational experiences, as well as a decrement in performance as completion of schooling becomes more distant. To begin to characterize the directionality of the relationship between educational attainment and cognitive performance, we conducted a prospective longitudinal analysis. For a subset of 69,202 subscribers who had completed 100 days of cognitive training, we tested whether the degree of novel learning was associated with their level of education. Higher educational attainment predicted bigger gains, but the differences were small ($d = 0.04$-$0.37$). Altogether, these results point to the long-lasting trace of an effect of prior cognitive challenges but suggest that new learning opportunities can reduce performance gaps related to one's educational history.

**Keywords:** educational attainment, executive function, plasticity, reasoning, college

# Introduction

Across industrialized nations, only a minority of adults complete post-secondary education. For example, fewer than 40% of adults in the United States are expected to graduate from college in their lifetimes, and the percentage shrinks for more advanced degrees (OECD, 2016). Considering the cost of higher education, many wonder whether it is a worthwhile investment. Nevertheless, post-secondary educational attainment has been consistently linked to financial and non-monetary benefits (Barrow & Malamud, 2015). Higher education is intended to confer the qualifications needed for the workforce, but also to improve individuals' critical thinking and readiness towards life-long learning (Harvey, 2000).

Indeed, universities may offer enriching experiences that enhance domain-general abilities to think and learn, such as thinking quickly (processing speed), keeping information in mind (working memory), responding flexibly to task goals (cognitive control), and tackling novel problems (reasoning). Although these skills are not explicitly taught in school, they may serve as a scaffold for learning and have been implicated in academic performance (Cattell, 1971; Deary, Strand, Smith, & Fernandes, 2007).

Prior research suggests that education has a positive effect on measures of intelligence (Ceci, 1991). For instance, longitudinal studies using data from compulsory military service in Scandinavian countries have estimated that each completed year of secondary school translates into a gain of nearly two to four IQ points during adolescence (Härnqvist, 1968) and early adulthood (Falch & Sandgren Massih, 2011). Moreover, the effects of schooling might be strongest for lower-performing individuals (Hansen, Heckman, & Mullen, 2003). The benefits of schooling have also been shown to be present in old adulthood, such that years of education predict IQ performance at the age of 70, even when controlling for individual differences in IQ at age 11 and other family characteristics such as parental socioeconomic status (Ritchie, Bates, Der, Starr, & Deary, 2013).

Analysis of data collected during periods of significant educational reform has provided even stronger evidence for the causal role of schooling on IQ. The most notable example is data analyzed by Brinch and Galloway (Brinch & Galloway, 2012), which spans nearly two decades when the Norwegian government raised compulsory schooling from seven to 9 years. The authors exploited the fact that different municipalities adopted the reform at various times and that men take a mandatory IQ test on the entrance to military service at age 18. These factors allowed the authors to compare the IQ of individuals who were able to leave school earlier than others. The analysis of this data indicated on average a benefit of nearly four IQ points for each year of schooling.

Aside from the composite measures of IQ used in the longitudinal and quasi-experimental studies described thus far, the effects of schooling have also been reported in studies that examine performance separately on tests that include scholastic knowledge and more abstract tests of cognition. Education positively predicts performance on the subcomponents of a typical IQ test including the reasoning and verbal portion (Cliffordson & Gustafsson, 2008; Kaufman, Kaufman, Liu, & Johnson, 2009), and it is a stronger predictor of performance on tests that directly measure skills taught in school, such as math and reading (Kaufman et al., 2009). The length of schooling has also been shown to positively predict performance on a measure of cognitive control in adolescence (Noble, Korgaonkar, Grieve, & Brickman, 2013) and measures of reasoning and working memory, but not processing speed, in old adulthood (Ritchie, Bates, & Deary, 2015). These findings suggest that educational experience has differential moderating effects on different aspects of cognition.

This prior body of work supports the notion that education positively influences higher cognition, consistent with principles of experience-dependent brain plasticity, from which one would predict improvements in cognitive skills that are repeatedly taxed in demanding and cognitively engaging coursework. However, the scope of prior work limits the conclusions that can be drawn, because they have focused on 1) limited cognitive domains or narrow age groups, 2) are mostly based on Scandinavian men who enlisted in the military, 3) lack the power needed to test the effects of different school levels or adequately characterize the effects of education across the life span, and/or 4) have not examined the impact of education on future learning. Here, we seek to build on prior work by addressing each of these issues.

### The present study

The goal of this study is to better understand the cognitive effects of education by testing whether educational attainment relates to cognitive abilities at one timepoint (a retrospective longitudinal approach), as well as learning efficacy from one timepoint to another (a prospective longitudinal approach). To this end, we examined performance on eight cognitive assessments of executive functioning and reasoning in a diverse sample of over 195,000 individuals, ages 15-60, who had subscribed to an online cognitive training program. Over 69,000 of these subscribers completed these assessments a second time approximately 100 days later, making it possible to evaluate practice-related gains in cognitive performance. We controlled for many variables, including income, sex, ethnicity, native language, and engagement with the training. Importantly, the engagement measures allowed us to quantify and control for individual differences in motivation in our learning context.

We hypothesize that if the cognitive assessments used here capture skills that are relevant to real-world outcomes, we should detect differences in performance associated with educational attainment (Ceci, 1991) in addition to age (Cliffordson & Gustafsson, 2008; Hartshorne & Germine, 2015). Given previous findings showing positive cognitive outcomes associated with continuing education in adolescence/young adulthood (Brinch & Galloway, 2012; Falch & Sandgren Massih, 2011; Härnqvist, 1968; Noble et al., 2013), we predicted a significant benefit of completing high school relative to not finishing it, and a further benefit of completing college. We also considered it plausible that there might be differences between holders of graduate degrees relative to college degrees, given differences in the years of higher education required.

Although we sought to understand the influence of education on cognition, it is incontrovertible that cognitive functioning itself influences educational attainment (Ferrer & McArdle, 2004). Some have argued that financial constraints, and not intellectual potential, are the major roadblock in educational attainment (Cardak & Ryan, 2009), but a selection bias is still to be expected, such that students with greater scholastic aptitude are more likely to pursue and attain higher degrees. While we cannot solve this chicken-and-egg problem short of randomly assigning students to pursue different degrees, the analyses described below help to address the question of how, and the degree to which, cognitive performance and learning efficacy vary as a function of prior education.

First, we examined whether educational attainment modulates age-related changes in cognitive performance at one timepoint. Previous work examining changes in cognition through the lifespan has shown that performance on some of the cognitive skills tested here peaks in late adolescence or early adulthood and declines thereafter (Hartshorne & Germine, 2015; McArdle, Ferrer-Caja, Hamagami, & Woodcock, 2002). Given the size and wide age range of our sample, it was possible to test whether these age effects are influenced by education – and, importantly, to determine how the cognitive effects of educational attainment differ across the lifespan, as one's experience with formal education recedes into the past and is supplanted by other life experiences. To this end, we explored whether educational attainment modulates ages of peak cognitive performance, such that the age of maximal cognitive performance for participants who have achieved a given level of education varies as a function of the age at which this degree is typically completed. We expected to replicate findings showing that late adolescence and early adulthood are the periods during which performance peaks for comparable measures of cognition (Hartshorne & Germine, 2015; McArdle et al., 2002). However, we further hypothesized that maximal cognitive performance would coincide with or closely follow the age at which education was completed. As an initial test of this hypothesis, we compared, for each educational level, the average age of peak cognitive functioning with the average age of graduation.

Second, we examined whether educational attainment differentially modulates performance on the eight individual cognitive assessments. Based on prior work in old adulthood showing the differential effect of schooling on various cognitive measures (Ritchie et al., 2015), we predicted that educational attainment would moderate age-related effects on tests of higher-level cognition, such as measures of reasoning, to a greater degree than on tests of lower-level cognition, such as measures of processing speed.

Finally, we examined prospectively whether educational attainment modulates learning efficacy. It has been argued that the effects of education are cumulative, such that quantity of schooling influence the acquisition and maintenance of cognitive skills over time (Ceci, 1991; Ganguli et al., 2010). To date, however, there is scant evidence for or against this hypothesis. Here, we sought to test whether prior education modulates practice-related gains in cognitive performance. To this end, we analyzed data from the subset of participants (*n* = 69, 202) who had completed the cognitive assessments before and after engaging with a cognitive training program. We considered three equally plausible outcomes. First, findings in the cognitive training literature (Diamond & Ling, 2016) raise the possibility that people starting with lower scores would improve the most. Conversely, considering the proposed cumulative effect of education and the well-documented Matthew effect (Walberg & Tsai, 1983), another possibility is that higher levels of education would predict greater gains. Lastly, given that the training games and assessments are unrelated to educational curricula, a third possibility is that there would be no effect of education on the magnitude of practice-related improvements.

To summarize, we examined retrospectively how educational attainment relates to cognitive performance in a large sample spanning the ages of 15-60. We examined the variance captured by educational attainment across the entire age range, and characterized how education moderates performance on individual cognitive assessments and well-established age-related changes in performance (i.e., ages of peak cognitive functioning). Finally, with our prospective analysis, we examined the effect of educational attainment on changes in cognitive performance before and after participation in a cognitive training program that taxes various aspects of executive functioning and reasoning.

## Methods

### Participants

Data were collected from Lumosity subscribers who had answered demographic questions and completed an online battery of cognitive assessments at least once. Subscribers were informed in advance that their data would be used for research purposes should they choose to complete the assessments. All data were de-identified and analyzed in aggregate in accordance with Lumos Labs' Privacy Policy (www.lumosity.com/legal/privacy_policy).

To be included in this study, participants had to be between the ages of 15-60 and reside in the United States (*n* = 152,694), Canada (*n* = 21,767), or Australia (*n* = 21,927). These countries were selected for several reasons. First, they were the most represented countries in the sample. Second, they share the same official language. Third, education is typically compulsory until the age of 16. Finally, their university systems require equivalent qualifications for admissions into college and graduate programs.

Additionally, we only included participants whose reported age was greater than their years of education (i.e., age > years of education + 4), and whose educational attainment was plausible given both their age and years of schooling (i.e., excluded participants younger than 17 with a Bachelor's degree or younger than 20 with a graduate degree). Moreover, we recoded participants' educational attainment as "Associate's/Some College" if they reported having a Bachelor's degree but were younger than 20 years old and had less than 15 years of formal education, since their response most likely reflects a clerical error rather than their actual educational attainment (*n* = 181). We chose these cutoffs based on pertinent international statistics of typical graduation ages (OECD, 2016). Finally, we also included in our analyses data from participants who did not specify their educational attainment level, so as not to bias the normalization procedures of the cognitive assessments scores.

Thus, the retrospective single timepoint analyses included 196,388 participants (53.72% females; $M_{\text{age}}$ = 39.95 ±12.8 *SD*). The prospective learning analyses included a subset of these participants (*n* = 69,202; 58.84% females; $M_{\text{age}}$ = 43.11 ±12.23 *SD*). The participants were from diverse demographic backgrounds and with educational attainment ranging from some high school to doctoral degrees (Fig 1, skewness$_{\text{years of education}}$ = -0.23). Our sample is slightly skewed towards higher educational attainment given that 52% of our sample has attained at least a Bachelor's degree, vs. ~34% of adults in the countries included in our analysis (OECD, 2016). However, the distributions of education categories across the age (Fig S1), income levels (Fig S2), and ethnic categories (Fig S3) are consistent with patterns seen in these general populations (OECD, 2016; Ryan & Bauman, 2016).

In addition to establishing education categories, we subdivided our participant into five-year age bins (e.g., 30-35) to maximize our ability to compare similarly represented educational attainment levels and age groups. We subdivided the age range into one-year bins to test whether the age of peak cognitive performance varies as a function of educational history.

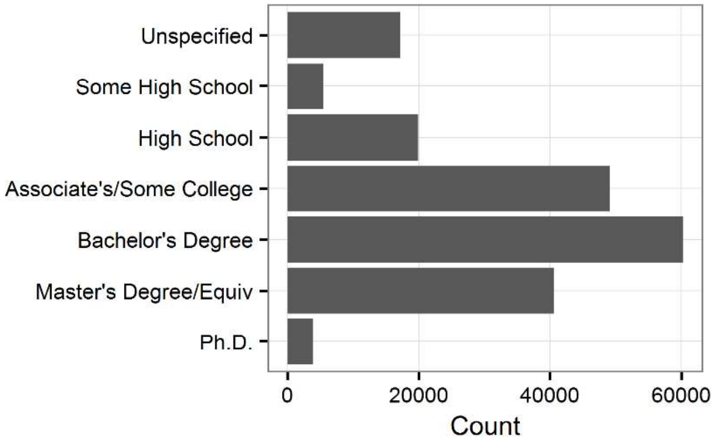


**Figure 1**. **Distribution of educational attainment**. The number of participants reporting educational attainment between Some High School to Ph.D.'s at T1 (*N* = 196,388).

### Cognitive Assessments

Participants completed a battery of eight assessments designed to evaluate working memory, flexibility, processing speed, and verbal and non-verbal reasoning. The assessments are a valid and reliable (Morrison, Simone, Ng, & Hardy, 2015) computerized adaptation of classic pencil-paper neuropsychological tests accessible online that can be completed in 30 minutes. These tests, including their performance metrics, have been described in detail elsewhere (Morrison et al., 2015). The working memory tests require participants to hold in mind strings of spatial locations (Forward and Reverse Spatial Span). The processing speed tasks involve quickly connecting numbers in a sequence (Trail Making A) and matching numbers to symbols (Digit Symbol Coding). The test of cognitive flexibility includes connecting in order-interleaved numbers and letters (Trail Making B). Finally, the tests of reasoning involve answering questions about simple logical statements (Grammatical Reasoning), performing simple word-based additions and subtractions (Arithmetic Reasoning), and completing a visual pattern based on rules and relationship between items (Progressive Matrices).

Participants completed these assessments at two timepoints, before (T1) and after (T2) engaging with the cognitive training program. We used T1 data for our retrospective analyses and the change score data (T2-T1) for our prospective analyses.

### Performance scores and analysis

The cognitive assessments have different performance metrics and distributions. Thus, we standardized the raw scores according to a conventional normalization procedure so that the performance scores will have a normal distribution with mean 100 (15 *SD*). Specifically, raw scores for a given subtest from T1 were ranked and then converted into percentile scores according to the empirical cumulative distribution. Normed scores were created by converting the percentile scores to their corresponding position on a normal distribution (*M* = 100, *SD* = 15).

We performed a similar normalization procedure for T2 scores, but they were converted to percentiles according to the empirical cumulative distribution of T1 raw scores to preserve changes in performance from T1 to T2. We used this normalization procedure to generate age-specific normative data.

We created an aggregate measure of T1 and T2 cognitive performance for each participant – a Grand Index (GI) score – by summing together for each timepoint the norm scores of each assessment and normalizing the sum scores to have a distribution with a mean of 100 (15 *SD*). Before testing for cognitive effects of educational attainment, we regressed out from all normed performance metrics the effect of demographic variables and engagement with the Lumosity games. We calculated GI change scores by subtracting the raw GI scores of T1 from T2. Before testing the effects of educational attainment, we regressed out from the change score demographic and engagement variables, as well as the raw T1 score.

Demographic covariates included participants' reported gender, ethnicity, and whether or not English was their native language. We included gender and ethnicity as covariates to control for differences in access to education and any effects of stereotype threat on cognitive test-taking ability (Steele, 1997). We also assessed effects of household income, given that income could influence access to education and other cognitively engaging activities. The engagement covariates were the number of hours each participant played the cognitive training games before each assessment and the number of days that elapsed between timepoints. We used the latter covariate only in the calculation of the change score. In all analyses, we used the log form of these engagement variables given their distribution and relationship to cognitive performance. Together, all these covariates accounted for a small variance in cognitive performance at both timepoints ($R^2 \leq 0.05, p < 0.0001$; Table S1).

All analyses use the adjusted normed scores resulting from these normalization procedures. The linear models used to probe the relationship of educational attainment with cognitive performance and learning are described in detail in the results section.

**Peak analysis procedures**

To test whether educational attainment modulates the ages of peak cognitive performance, we adopted a bootstrap resampling procedure similar to one previously employed with large cross-sectional online samples (Hartshorne & Germine, 2015). Specifically, we drew a sample between ages of 15-60 (in one-year bins) from each educational category and identified the age group with the highest T1 GI score, using the age-specific normative data and adjusting for all the aforementioned control variables. We repeated the procedure 10,000 times, which allowed us to calculate a median age of maximal performance and the corresponding 95% confidence intervals (CIs). The sample size selected (with replacements) from each education group at each iteration equaled the sample size available for that education category (Fig 1), and we only considered age bins with at least 50 participants for each education level.

To explore the possibility that age at peak cognitive functioning is related to the age at which education is terminated, we accessed international indicators of typical ages of graduation for each educational attainment category (OECD, 2016). The report listed ages of graduation as a range separately for each country included in our sample, which we used to calculate the median age and typical age ranges of graduation for each education category. For instance, we estimated that for a Bachelor's degree, the typical ages of graduation are 20 to 24 and the median age is 22.5 given the typical ages of graduation in each country: 21-23 in the USA, 22-24 in Canada, and 20-23 in Australia.

**Supplementary materials**: supplementary information referenced in this chapter can be found in Appendix A.

## Results

### Effects of educational attainment and age on cognitive performance

To test whether educational history modulated cognitive performance across a range of task demands, we used the aggregate measure based on all the cognitive assessments for each participant at T1, termed the Grand Index (GI) score. As described above, we controlled for demographic variables and level of prior engagement with the Lumosity training games.

As predicted, educational attainment levels (Some High School, High School, Some College/Associate's, Bachelor's, Master's, Ph.D.) and age (categorical 5-year age bins between ages 15 and 60), were significant predictors of the *adjusted* GI score at T1 (*adjusted* $R^2 = 0.18$; $p < 0.0001$; Table 1; Fig 2). Household income did not account for additional variance (*adjusted* $R^2 = 0.18$, $p < 0.0001$; Table 1) and was therefore not included in subsequent analysis.

**Table 1. Educational attainment and age predicting T1 cognitive performance and learning, as measured by the Grand Index (GI) score.**

| | | Controlling for income | | Controlling for T1 scores | Age regressed out | |
|---|---|---|---|---|---|---|
| | **GI T1** Estimate | **GI T1** Estimate | **Δ GI** Estimate | **Δ GI** Estimate | **GI T1** Estimate | **Δ GI** Estimate |
| **Intercept** | 100.63*** | 100.87*** | 5.12** | 5.20 *** | 95.11 *** | 2.95*** |
| **Some high school** | -2.73*** | -2.69*** | 0.05 | -0.33 | -1.80 *** | -0.31 |
| **Assoc./Some College** | 3.90*** | 3.90*** | 0.31** | 0.69*** | 3.10 *** | 0.73*** |
| **Bachelor's** | 7.83*** | 7.78*** | 0.42*** | 1.20 *** | 6.63 *** | 1.31*** |
| **Master's/equivalent** | 8.91*** | 8.83*** | 0.52*** | 1.42 *** | 7.65 *** | 1.58*** |
| **Ph.D.** | 10.81*** | 10.74*** | 1.12*** | 2.20 *** | 9.64 *** | 2.44*** |
| **Unspecified** | 3.43*** | 3.46*** | 0.36* | 0.78*** | 2.50 *** | 0.76*** |
| **Ages 20-25** | 0.52** | 0.39** | -0.43* | -0.27 | | |
| **Ages 25-30** | -0.39* | -0.58* | -0.70*** | -0.63** | | |
| **Ages 30-35** | -2.14*** | -2.36*** | -0.98*** | -1.05 *** | | |
| **Ages 35-40** | -3.86*** | -4.10*** | -1.42*** | -1.67 *** | | |
| **Ages 40-45** | -5.97*** | -6.22*** | -1.81*** | -2.27 *** | | |
| **Ages 45-50** | -8.59*** | -8.85*** | -1.73*** | -2.49 *** | | |
| **Ages 50-55** | -11.99*** | -12.23*** | -1.73*** | -2.86 *** | | |
| **Ages 55-60** | -15.11*** | -15.33*** | -2.07*** | -3.51 *** | | |

The GI change score (ΔGI) is the difference between the GI score from T1 and T2. Each GI score was normalized to have a distribution with mean of 100 (15 SD) and was adjusted for the effects of demographic covariates (gender, ethnicity, and indicator of English as native language), engagement variables (number of gameplay hours and days between T1 and T2), T1 performance (T2 only), and other specified variable. The two rightmost columns additionally regress out the effects of age (5-age bins). Reference category: ages 15-20 and High School attainment. $p < 0.0001$'***', $p < 0.001$'**', $p < 0.01$'*'

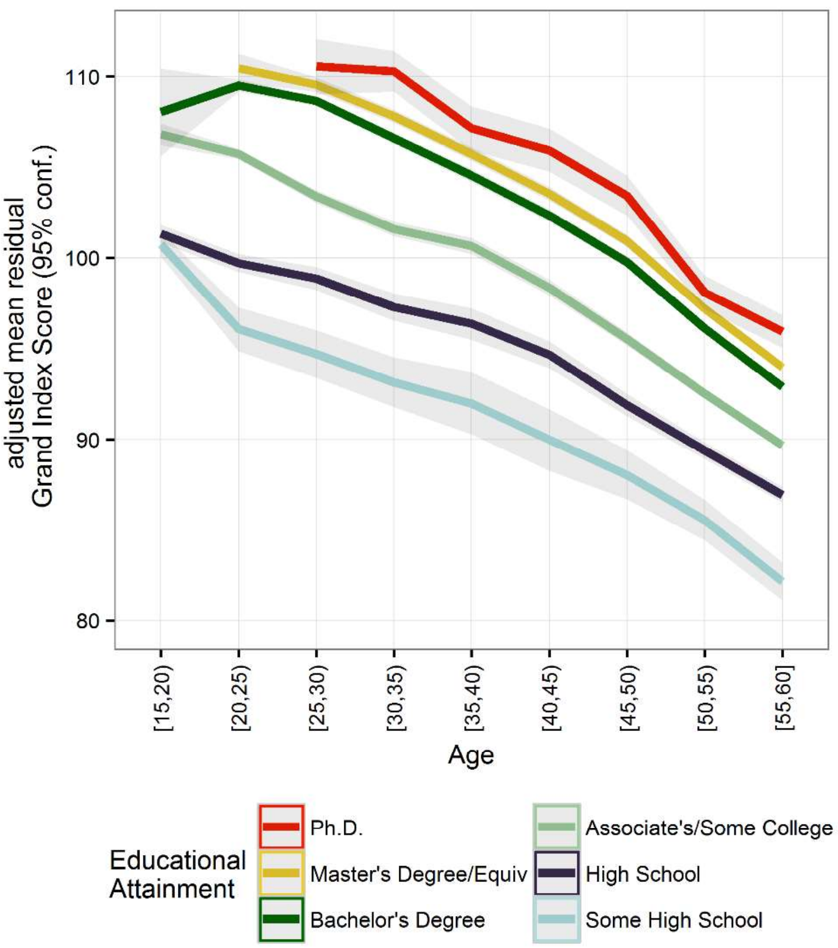


**Figure 2**. **Effect of educational attainment on the adjusted Grand Index score at T1** across the ages of 15 and 60. Ribbons show bootstrapped 95% CIs from 10,000 iterations. Displaying age/education categories with 50+ participants who specified their educational attainment level (*n* = 179,141; *N* = 196,388).

## Quantifying the unique influence of educational attainment

Having found that GI scores vary as a function of both age and education, we sought to quantify the amount of unique variance explained by the latter. To this end, we regressed out the 5-year age bins in addition to the demographic and engagement covariates when calculating the GI score. A linear regression showed that educational attainment positively predicted a significant, albeit small, amount of variance in cognitive performance (*adjusted* $R^2 = 0.04$, $p < 0.0001$). All educational attainment categories were significant predictors of the GI score at T1, showing a pattern of results in line with the additional years of schooling required to complete each degree (Table 1).

Next, we sought to more carefully characterize the effects of different levels of educational attainment. Thus, we calculated pairwise differences in effect size between each educational attainment levels using Cohen's d (Table 2). Differences in effect sizes were relatively large between the extremes of educational attainment (Ph.D. vs. Some High School, *d* = 0.80), moderate at a key educational juncture (Bachelor's vs. High School, *d* = 0.51), and small between other adjacent education levels (e.g., Master's vs. Bachelor's, *d* = 0.08).

**Table 2**. Pairwise differences in effect size between educational attainment levels at T1

| | **Some High School** | **High School** | **Some College/ Associates** | **Bachelor's** | **Master's or equivalent** |
|---|---|---|---|---|---|
| **High School** | 0.13 * | - | | | |
| **Some College/ Associates** | 0.36 * | 0.23 * | - | | |
| **Bachelor's** | 0.65 * | 0.51 * | 0.27 * | - | |
| **Master's or equivalent** | 0.72 * | 0.59 * | 0.35 * | 0.08 * | - |
| **Ph.D.** | 0.80 * | 0.71 * | 0.49 * | 0.23 * | 0.16 * |

Pairwise effect sizes (Cohen's d) were calculated between each educational attainment category predicting T1 Grand Index Score with age regressed out. CIs calculated using 10,000 bootstrap iterations. '*' 95% CIs does not include 0.

### Educational attainment moderates the age of peak cognitive performance

Having found that education and age are stronger predictors of performance than education alone, we performed a granular analysis examining the effects of educational attainment on cognitive performance across the lifespan. Specifically, we tested whether the age of maximum performance differed between education categories. The literature suggests that late adolescence/early adulthood is when performance peaks on comparable measures of cognition (Hartshorne & Germine, 2015; McArdle et al., 2002). We reasoned that if educational attainment influences cognitive performance and these effects are greater as the educational experience are more recent, the age of peak performance would vary as a function of education and be proximal to the ages when people typically complete the education programs. We found that ages of peak performance were within the young adulthood period, occurred later the higher the education level, and overlapped with typical ages of graduation for each degree (Fig 3). Specifically, 17 was when performance was maximum for the High School and age 22 for the Bachelor's category. These ages were well-aligned with the typical ages of graduation from those educational categories, 17.5 and 22.5 respectively (OECD, 2016). Ages of peak performance for the other degrees were within the range of typical ages of graduation for their respective program. Importantly, peak performance was not at the youngest possible ages for each education category, which suggests that we may not be just capturing an effect of age or performance from higher achieving individuals (i.e., youngest people to receive a postsecondary degree), but instead the effects of a recent educational experience.

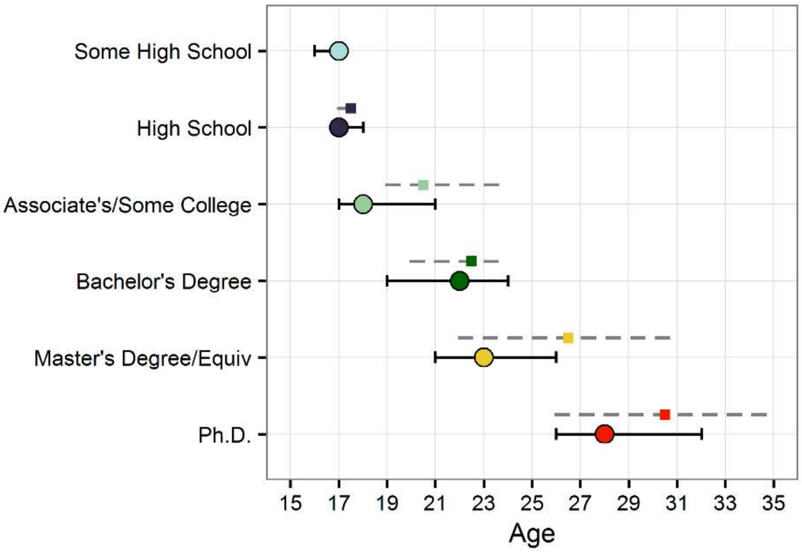


**Figure 3**. **Educational attainment moderates the ages when cognitive performance peaks**. Colored points show the median age of maximum performance and the error bars the 95% CIs. We calculated these ages using a 10,000 iteration bootstrap sampling procedure including the entire age range available in our sample (ages 15-60). Gray dotted lines represent the age range, and the colored squares represent the median age of typical graduation for each education level. We obtained these graduation ages from international indicator reports, which included data from all three countries represented in our sample (USA, Canada, and Australia).

## Influence of educational attainment on individual cognitive measures

Thus far, we have presented how education relates to cognitive performance on an aggregate measure (GI) because it is the most reliable and robust measure (Morrison et al., 2015). However, we also sought to explore the possibility that educational attainment was specifically or preferentially related to a subset of the eight cognitive assessments, moderating typical age-related changes in performance. Thus, we tested the effect of educational attainment and age on performance on tests of working memory, processing speed, cognitive flexibility, and verbal and non-verbal reasoning. The tests were free of educational content – except for Arithmetic Reasoning, which involved elementary scholastic content (e.g., simple word-based additions and subtractions) -- that all our participants are expected to have been exposed to in school. We hypothesized that higher education would have its largest effect on measures of reasoning, given the complex, abstract material covered in college and beyond, and that it would have its smallest effect on tests of processing speed.

As predicted, age and education predicted distinct amounts of variance in performance on different assessments (Fig 4). On simple tests like Trail Making A, which requires speeded responding based on a simple rule, there was a noticeable age-related decline and a weaker effect of education. By contrast, scores on more cognitively complex tests, including Arithmetic Reasoning, Grammatical Reasoning, and Progressive Matrices, showed greater variance explained by educational attainment when controlling for age (Fig S4). These assessments also tended to show later peak performance and initial points of decline as a function of age.

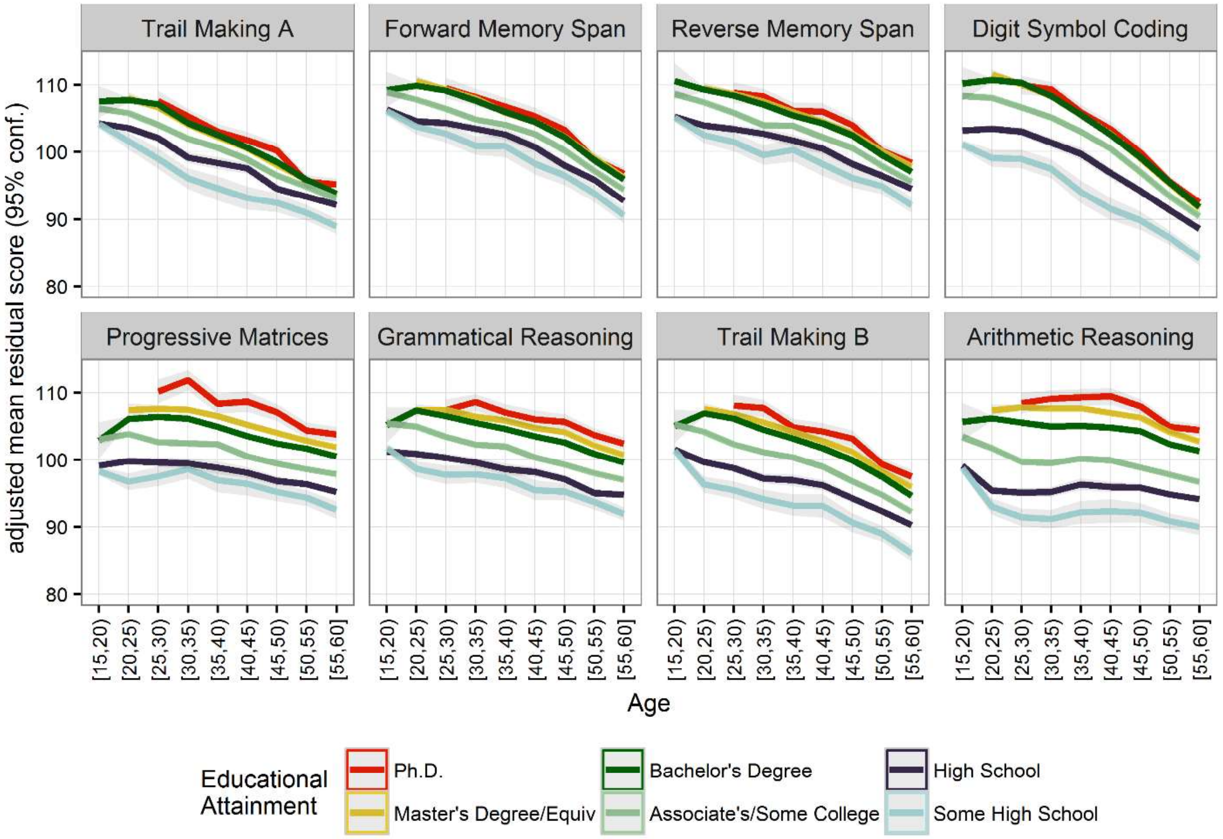


**Figure 4. Effects of educational attainment on individual cognitive assessments across ages 15-60** ($n$ = 179,141). Ribbons show bootstrapped 95% CIs based on 10,000 iterations. Displaying age/education categories with 50+ participants who indicated an educational attainment level ($N$ = 196,388).

Considering that the test that was most strongly modulated by education, Arithmetic Reasoning, required numerical competencies (albeit skills taught in elementary school), we tested whether our results would hold when excluding this assessment. Indeed, excluding it from the calculation of the GI score at T1 did not impact the amount of variance educational attainment explained, nor the differences in effect sizes between education levels. For example, educational attainment and age (categorical 5-year age bins between ages 15 and 60) remained significant predictors of the adjusted GI score (*adjusted* $R^2 = 0.19$; $p < 0.0001$). Thus, educational history modulated performance even on tasks that bore no resemblance to those encountered in the classroom.

## Effects of educational attainment and age on learning

To test whether prior educational attainment influences how quickly one learns, we took advantage of the fact that 69,202 of the subscribers in the sample took the cognitive assessments twice, on average 100.77 days (± 57.67 *SD*) apart, playing a suite of cognitive games in the interim (*M*= 166.23 hours ± 290.20h *SD*). These games were different from the assessments but were designed to tax the same underlying cognitive skills (Hardy et al., 2015). The goal of this study was not to assess the overall effectiveness of the training (Hardy et al., 2015), but rather to test whether educational attainment would be associated with the magnitude of gain in the GI score. Thus, we calculated a GI change score as the difference in performance between the assessments taken after (T2) and before the training program (T1). We then tested whether educational attainment modulated the GI change score, regressing out the effects of the same demographic variables as T1, engagement in the training program (i.e., number of hours of gameplay and days elapsed between assessments), and T1 performance (to control for the plausible effect of regression to the mean).

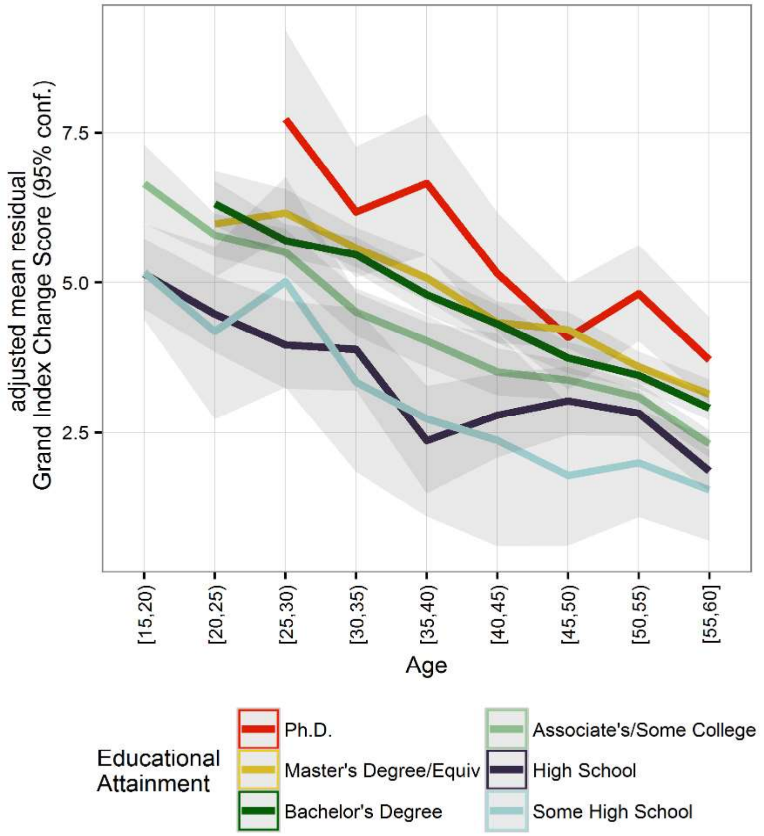


**Figure 5. Effect of educational attainment on adjusted GI change score across ages 15-65** (*n* = 63,535). Y-axis units represent the change in GI scores from T1 to T2. At each timepoint, the GI scores were normalized to have a distribution with a mean of 100 (15 *SD*). Ribbons show bootstrapped 95% CIs based on 10,000 iterations. Displaying age/education categories with 50+ participants who indicated an educational attainment level (*N* = 69,202).

## Quantifying the unique influence of educational attainment on learning

We next sought to quantify the effect of educational attainment on the change in cognitive performance independently of the effect of age. To this end, we calculated the GI change score with the 5-year age bins as additional covariates, following similar procedures to the analogous T1 analysis. Educational attainment accounted for a negligible amount of variance in the adjusted GI change score (*adjusted* $R^2$= 0.01, $p < 0.0001$; Table 1). Post-secondary attainment predicted additional improvements from the High School reference category, but overall, participants across education levels exhibited similar gains. The fact that educational attainment was not negatively related to the change score leads us to reject the hypothesis that individuals that are more educated stood to gain less from practice than their peers because of their higher starting performance level.

We also compared the effects of different levels of educational attainment with pairwise differences in effect size between each education category (Table 3). As with T1 data, differences in effect sizes for the magnitude of learning were larger between the extremes of educational attainment (Ph.D. vs. Some High School, $d = 0.37$) than between adjacent levels (e.g., Master's vs. Bachelor's, $d = 0.04$). All of the results above of the effects of educational attainment on learning were replicated after excluding the Arithmetic Reasoning test from the calculation of the GI score at both timepoints. Thus, as with cognitive performance at one timepoint, practice-related gains were associated with educational attainment even for non-academically related tasks.

**Table 3**. **Pairwise differences in effect size between educational attainment levels in the change score analysis**

| | Some High School | High School | Some College/ Associates | Bachelor's | Master's or equivalent |
|---|---|---|---|---|---|
| **High School** | 0.04 | - | | | |
| **Some College/ Associates** | 0.14 * | 0.10 * | - | | |
| **Bachelor's** | 0.22 * | 0.18 * | 0.08 * | - | |
| **Master's or equivalent** | 0.26 * | 0.22 * | 0.12 * | 0.04 * | - |
| **Ph.D.** | 0.37 * | 0.33 * | 0.24 * | 0.16 * | 0.12 * |

Pairwise effect sizes (Cohen’s d) were calculated between each educational attainment category predicting Grand Index Change Score with age regressed out. CIs calculated using 10,000 bootstrap iterations. ‘*’ 95% CIs does not include 0.

## Discussion

We sought to test the extent to which cognitive performance and learning efficacy in adolescence and adulthood vary as a function of educational attainment, ranging from some high school to advanced degrees. To this end, we analyzed data from a large sample of 15-60-year-olds who completed eight cognitive assessments. Controlling for multiple demographic and engagement variables, we found that educational attainment had a small but significant overall effect on performance at the initial timepoint, but had a negligible effect on learning efficacy. There were moderate differences in performance between secondary and post-secondary education levels, but minimal ones between post-secondary levels.

Peak analyses revealed that the higher the education level, the later the age of maximal cognitive functioning. Further, the peaks overlapped with typical graduation ages for the different degrees. This overlap was tighter for the High School and Bachelor's degree, for which the age of maximal performance coincided almost exactly with typical ages of graduation. For the other degrees, the age of peak performance was within the range of typical graduation ages. These results suggest that we may be capturing the cumulative effect of a recent educational experience. Although these preliminary results are cross-sectional and the dataset did not include the age at which individual participants completed their education, they are suggestive of an age-related cognitive decline beginning shortly thereafter.

The variance in performance explained by education was greater in some cognitive domains, such as reasoning, but nonetheless smaller compared to the effects of age. However, the effect of education was present across the broad age range, persisting for decades beyond typical graduation ages. Although we did not have a principled way of rank-ordering or grouping the eight assessments according to their level of cognitive complexity, our results are suggestive of the idea that education affects higher-level cognitive functions more strongly than lower-level ones. This observation complements previous findings (Ritchie et al., 2015) by documenting the effect across the five decades of life spanned by our sample.

We found a modest effect of educational attainment on learning, as indexed by gains on the cognitive assessments after completing the training program. Post-secondary education categories exhibited only slightly larger gains than secondary levels. Moreover, High School graduates reached scores at post-test that were comparable to those attained at pre-test by individuals who had completed some college. Thus, practice may reduce gaps in performance observed as a function of educational history. The fact that education had only a small effect on learning is perhaps not surprising, given that the training program was not academic in nature. The cumulative benefits of education may be more salient when the curricula build directly on academic knowledge and skills explicitly taught in school.

## Limitations and Future Directions

Our study has several limitations. First, it is impossible with our cross-sectional dataset – or even with our dataset including two timepoints of data per individual – to prove conclusively that higher education hones domain-general cognitive functioning. In fact, conclusively teasing apart the bi-directional influence of cognition and education is insurmountable even when individuals are followed for years because having an initial assessment of cognitive performance before undergoing an educational experience does not preclude the effect of other confounding variables (e.g., motivation) (Hansen et al., 2003). Data from "natural experiments," such as instances of school reform, can provide stronger evidence on the directionality of effects, but these situations are rare and may suffer from other confounding variables (Brinch & Galloway, 2012).

Despite this inherent challenge, the current study complements those that have employed a longitudinal design or a naturalistic experiment, since we demonstrate a positive relationship between educational attainment and cognitive performance within the range of effects in literature. Moreover, the unique size and heterogeneity of our sample allow us to quantify the effects of educational attainment relative to other factors, including the influence of demographic and engagement variables. Further, our analysis of age at peak cognitive functioning suggests that we are able to capture the effects of recent educational experiences on cognition and not just general effects of age. Also, the fact that our cognitive tests are so different from school curricula yet show differential effects of education helps further disambiguate the relationship between educational attainment and cognitive performance in our study. Finally, the test-retest data allowed us to test how prior educational experiences influence the efficacy of new learning.

A second limitation is that our analyses hinged on self-reports of age, education, income, and other demographic variables. Errors in self-reporting could have led to miscategorization of participants, even though we took steps to remove participants whose answers were incongruous. If anything, however, miscategorization would introduce noise into the dataset that would likely lead to an underestimation of the effect sizes. We also do not have a measure of the participant's childhood socioeconomic status or whether they completed their education in the countries included. Again here, however, such errors should reduce rather than inflate our ability to detect predicted effects. Relatedly, not knowing when our subjects completed their degrees, but instead inferring this information from reports of international indicators, constrains the interpretations that can be drawn from our analysis examining whether educational attainment modulates ages of peak cognitive performance. However, to maximize the possibility that our sample was representative of the larger population, we only included in the analysis age groups with a representative number of subjects in each education category.

Another potential limitation is that there is a possible selection bias for subscribing in Lumosity. However, the broad distribution of our sample's education across demographic variables is similar to patterns observed in the general population, and the age-related declines in performance are systematic and consistent with the literature. We could alternatively consider this potential limitation as a feature of the dataset. If there is, in fact, a selection bias for subscribing to Lumosity such that our sample represents individuals who are motivated to pursue cognitive enriching activities despite their educational history, then motivation-related confounds – that have been hypothesized to drive in part the relationship between education and cognitive performance (Hansen et al., 2003) – should be attenuated in this sample.

Finally, facility with computers could have contributed to our findings, given that the assessments were computerized and that prior experience with computers likely varied both as a function of age and education (OECD, 2016). However, our results are inconsistent with this account: for one, individual assessments were modulated differently by age or education; for another, if the results reflected computer skills or acquisition, the subset of participants expected to have better computer skills (e.g. higher education/younger ages) should benefit the least from the training program, not the most.

Our unique dataset allowed us to begin answering important questions about the cognitive effects of education that should be further investigated with longitudinal studies. For instance, performance differences between education levels were evident from our earliest 5-year age bin group. However, it remains an open question whether the magnitude of the education effects increase, narrow, or remain stable with age. Additionally, findings from the peak analysis showing that maximal performance and the subsequent decline occurs later in the higher education levels are consistent with the idea that higher education may help to stave off age-related cognitive decrements (Ganguli et al., 2010; Shimamura, Berry, Mangels, Rusting, & Jurica, 1995). These observations raise a question about whether or how the timing of the educational experiences impacts cognitive functioning over the long term. Finally, our results from the learning efficacy analysis are consistent with findings showing that young adults show larger gains than older adults from cognitive interventions (Simone Kühn & Lindenberger, 2016). The question remains, however, the degree to which prior educational experiences interact with the effects of age. The answers to these and similar questions have theorerical implications to increase our understanding of the sensitive periods in the development of higher cognition and its plasticity through adulthood, as well as practical implications for govermental decisions about school reform and policy geared towards increasing inclusive access to and completion of higher education.

In conclusion, our results indicate a relation between educational attainment and cognitive abilities across a broad age range but small effects on learning efficacy. Although our results are statistically significant, even after controlling for multiple potential confounds, the amount of variance explained by educational attainment is small. In a smaller-scale study, these effects would probably not have been detected at all (Deary et al., 2007). Nonetheless, these findings support the idea that higher education provides the opportunity to hone domain-general cognitive skills as well as to acquire content knowledge and that education-related gap in performance can be mitigated with intensive cognitive engagement.

**Next section, Chapter 4-6**

In Chapter 2 and 3 I provided evidence of the impact of learning in the brain and performance. Now, we turn our attention to probing mechanisms. In Chapter 4, I present unified findings from a broad literature to explain how various eyetracking measures constitute a relatively simple method for characterizing cognitive mechanisms of change. Using this methodology, in Chapter 5 and 6, I provide empirical evidence about the mechanisms that support improvements in reasoning performance in young adults who underwent a real-world learning experience.

# Chapter 4

## Beyond eye gaze: What else can eyetracking reveal about cognition and cognitive development?

Co-authors: Maria K. Eckstein[1], Alison T. Miller Singley[1], & Silvia A. Bunge[1,2]

[1]Department of Psychology & [2]Helen Wills Neuroscience Institute
University of California at Berkeley

**Permission to include co-authored material:** Granted

**Acknowledgements:** We thank everyone who helped us launch our eyetracking laboratory: Mehdi Bouhaddou, Lisa Johnson, Sheri Johnson, Jesse Niebaum, Andrew Peckham, Jordan Tharp, and Carter Wendelken. We also acknowledge the following funding sources: German Academic Exchange Service Doctoral Scholarship (M.E.), National Science Foundation Graduate Research Fellowship (B.G-C.), James S. McDonnell Foundation 21st Century Science Initiative Scholar Award in Understanding Human Cognition (S.A.B.), and a Jacobs Foundation Advanced Career Research Fellowship (S.A.B.).

## Abstract

This review provides an introduction to two eyetracking measures that can be used to study cognitive development and plasticity: pupil dilation and spontaneous blink rate. We begin by outlining the rich history of gaze analysis, which can reveal the current focus of attention as well as cognitive strategies. We then turn to the two lesser-utilized ocular measures. Pupil dilation is modulated by the brain's locus coeruleus-norepinephrine system, which controls physiological arousal and attention, and has been used as a measure of subjective task difficulty, mental effort, and neural gain. Spontaneous eyeblink rate correlates with levels of dopamine in the central nervous system, and can reveal processes underlying learning and goal-directed behavior. Taken together, gaze, pupil dilation, and blink rate are three non-invasive and complementary measures of cognition with high temporal resolution and well-understood neural foundations. Here we review the neural foundations of pupil dilation and blink rate, provide examples of their usage, describe analytic methods and methodological considerations, and discuss their potential for research on learning, cognitive development, and plasticity.

**Keywords**: eyetracking; saccades; pupillometry; pupil dilation; blink rate; children

**Highlights**:

- Eyetracking measures provide non-invasive and rich indices of brain function and cognition
- Gaze analysis reveals current attentional focus and cognitive strategies
- Pupil dilation is modulated by norepinephrine and reflects mental effort
- Spontaneous blink rate is modulated by dopamine, which is involved in learning and goal-oriented behavior
- Ocular measures can provide insights regarding cognition and cognitive development

# Introduction

A remarkable insight from the field of psychology is the fact that we can probe the inner workings of the mind by measuring how various eye muscles contract (**Figure 1**). Cognitive psychologists have exploited this fact for over two centuries (e.g., Wells, 1792; and Hering, 1879; cited by Wade, 2015). Over the last two decades, however, eyetracking has largely taken a backseat to brain imaging research as a way to study the mechanisms that underlie behavior. Now, thanks to notable improvements in eyetracking hardware, software, and analytic approaches, as well as increased recognition of the limits of what we can learn from brain imaging, eyetracking is regaining its former status. The overarching goals of this review paper are threefold: first, to provide an overview of ocular measures and what we have learned from studies in adults about their neurobiological underpinnings and behavioral correlates; second, to discuss methodological approaches and considerations; and third, to discuss how eyetracking has been and could be extended to study cognitive development. The most commonly utilized ocular measure is that of eye gaze; we will provide only a brief overview of this approach before focusing primarily on task-evoked pupillary responses and spontaneous eyeblink rate. This review focuses on the applicability of these measures to our understanding of cognitive functioning in neurotypical children and adults; however, this methodology is also useful in clinical research (Blaser, Eglington, Carter, & Kaldy, 2014; Burkhouse, Siegle, Woody, Kudinova, & Gibb, 2015; Caplan & Guthrie, 1994; Chan & Chen, 2004; Fried et al., 2014; M. Hallett, 2000; C. N. Karson, 1988; Rommelse, Van der Stigchel, & Sergeant, 2008; Tulen et al., 1999).

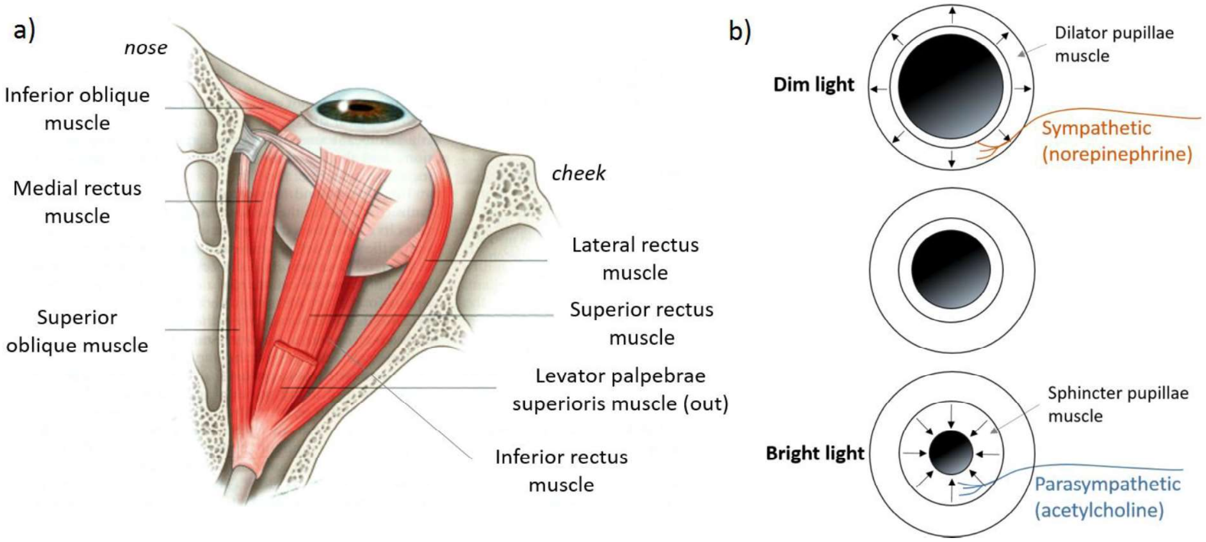


**Figure 1.** Eye muscles responsible for eye movements and pupil dilation and contraction. a) Superior view of the eye. The superior and inferior rectus muscles are responsible for the eye's vertical movements, whereas the lateral and medial rectus muscles control horizontal movements. Adapted with permission from Eds. Levin, Nilsson, Ver Hoeve, Wu, Kaufman, and Alm (2011). b) Top: The dilator pupillae muscle dilates the pupil and is controlled by sympathetic fibers. Bottom: The sphincter pupillae muscle contracts the pupil and is controlled by parasympathetic fibers. The balance between the activation of the dilator and sphincter pupillae muscles dictates pupil diameters.

One might think of eyetracking as either an impoverished measure of brain function or a rich measure of cognition. However, it can complement both behavioral and brain measures. Indeed, it has been argued that oculomotor studies provide an ideal neuroscience model to investigate association between brain mechanisms and behavior (e.g., Luna et al., 2008). Ocular measures can provide additional information over and above accuracy and response times as a result of their high temporal resolution, making it possible to measure how people respond to task demands on a moment-by-moment basis. Indeed, eyetracking sampling rates range from 25-2000 measurements per second, which means that the faster eyetrackers achieve sub-millisecond temporal resolution, similar to EEG. Despite being an indirect measure of brain function, eyetracking has several advantages compared to EEG and fMRI, which make it the better choice for a number of paradigms and research questions. First, given that participants can be seated comfortably at a table during data collection (or can move freely, with a head-mounted eyetracker), testing can happen in a more natural environment than the noisy and space-restricted environment of the MRI scanner. Second, most eyetrackers are portable, making it possible to take them to schools, hospitals, and other venues. As such, it is possible to reach a larger and more diverse population than the small pool of participants who are willing and able to travel to research facilities. Third, the rapid calibration procedures available on modern eyetrackers make it possible to begin an experiment quickly. This is particularly helpful for developmental researchers seeking to minimize testing time.

In many studies, ocular data are captured for the sole purpose of ensuring that participants maintain fixation at the center of the screen. However, the measurement of eye position can also provide a moment-by-moment assessment of thought processes in a wide variety of contexts (e.g., Shepherd, Findlay, & Hockey, 1986; Theeuwes, Belopolsky, & Olivers, 2009; Van der Stigchel, Meeter, & Theeuwes, 2006). Yarbus (1967) provided a simple illustration of this idea, asking subjects different questions as they viewed the same painting. When asked to judge the age of each character, the sample participant looked primarily at the depicted faces; when asked to judge the material wealth of the family, he looked primarily at the characters' clothing and some of the surrounding objects.

In the decades since Yarbus' vivid demonstration, numerous studies have corroborated the fact that one's eyes are generally directed towards the object of one's thoughts (Ferreira, Apel, & Henderson, 2008; Just & Carpenter, 1976, 1980; Theeuwes et al., 2009; Thomas & Lleras, 2007; Van der Stigchel et al., 2006). Indeed, although it is possible to attend covertly to a spatial location without moving one's eyes to it, it is not only more common but also more effective to fixate what we are attending to (Deubel & Schneider, 1996; Shepherd et al., 1986; Theeuwes et al., 2009). Further, fMRI and cortical stimulation research corroborate the close link between attention and gaze, showing that the frontal eye fields (FEF), which control eye movements, are also implicated in the deployment of covert visual attention (Awh, Armstrong, & Moore, 2006; Corbetta et al., 1998; Grosbras, Laird, & Paus, 2005; Müller, Philiastides, & Newsome, 2005).

Multiple gaze metrics that have been used to study cognition in adults are derived from eye position data. Fixations are used to calculate time spent looking at a particular location, which in turn is thought to reflect engagement of attention and the time needed to process the stimulus at that location. This metric has been used to gain insights into what we remember

(Hannula et al., 2010), how we perform mental computations (e.g., Green, Lemaire, & Dufau, 2007), how we read (e.g., Rayner et al., 1998), how we solve problems (Grant & Spivey, 2003), and how we learn (Lai et al., 2013; Rehder & Hoffman, 2005). Saccades, the rapid eye movements that allow us to shift between fixations, can reflect shifts in attention that are either controlled (e.g. a voluntary eye movement or saccade towards a target) or automatic and stimulus-driven (e.g., a reflexive saccade towards a sudden stimulus) (Luna, Velanova, & Geier, 2008). The accuracy and latency of saccades have provided insights for example about cognitive control capacity (e.g., Funahashi, Bruce, & Goldman-Rakic, 1989; Luna & Velanova, 2011; Munoz & Everling, 2004). The number of saccades between task-relevant stimuli, which is assumed to reflect the process of comparing specific stimuli or integrating several pieces of information, has been used to study reasoning (Demarais & Cohen, 1998; Thibaut & French, 2016; Vigneau, Caissie, & Bors, 2006); Some research questions require analysis of scan paths, rather than simple quantification of fixation and saccade measures. In these types of investigations, the subject of interest is how people approach a problem space, so the measure must encapsulate multiple fixations and the path of movements between them (Bochynska & Laeng, 2015; Dewhurst et al., 2012; Hayes, Petrov, & Sederberg, 2011; Yoon & Narayanan, 2004).

We have identified three broad classes of eye movement studies of cognitive development (**Figure 2**). The first of these is comprised of studies that measure reflexive orienting to a stimulus (i.e., reflexive saccades). The second class involves tasks in which the target response is a voluntary eye movement (i.e., voluntary saccades), wherein measurement of saccades is needed to measure task accuracy and response latency. Finally, the third class involves measures of spontaneous eye gaze patterns (i.e., scan paths) during analysis of a complex stimulus or a set or series of stimuli, for example in studies of higher-order cognitive abilities like reading or reasoning.

The first class of studies has been particularly useful for studying cognitive processes in infancy, since the brain pathways that control reflexive saccades are relatively mature at birth, whereas those controlling voluntary eye movements are immature at birth but develop rapidly during the first six months of infancy (Richards & Hunter, 2002).

In an effort to understand what they know or remember, researchers measure how long infants look at a novel or unexpected stimulus. This implicit measure of attention is analogous to 'looking time' measures of head turns towards an object, but is considered more precise (Aslin & McMurray, 2004; Feng, 2011; Franchak, Kretch, Soska, & Adolph, 2011). Eye movement analyses have already been used extensively to characterize the expectations and cognitive processes of infants (for a review see Gredebäck et al., 2009).

In a study of language comprehension, for example, Lewkowicz and Hansen-Tift (2012) showed that 8-10-month-old infants looked longer at a speaker's mouth than at her eyes, whereas younger infants and older children and adults both showed the opposite preference (**Figure 2a**). All infants looked more at the speaker's mouth during infant-directed speech as opposed to adult-directed speech. The authors' interpretation is that infants direct their attention to the mouth as they learn to produce sounds, whereas older children and adults tend to focus on the eyes to glean social cues (Lewkowicz & Hansen-Tift, 2012). A separate longitudinal study showed that the transition from focusing on a speaker's eyes to her mouth happened between 6 and 9 months of age (Tenenbaum, Shah, Sobel, Malle, & Morgan, 2013). This approach has also been used to probe infants' sustained and joint attention, object representation, perceptual completion, and relational memory (Gredebäck, Johnson, & von Hofsten, 2009; Johnson, Amso, & Slemmer, 2003; Richmond & Nelson, 2009; Yu & Smith, 2016).

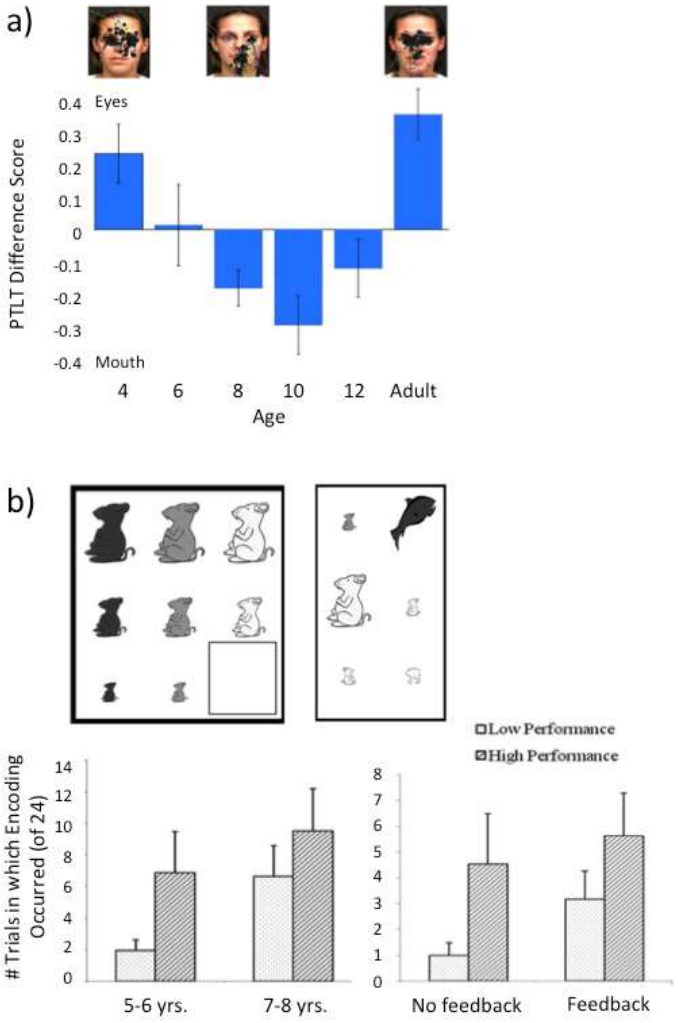


**Figure 2**. Gaze analyses in developmental research on attentional capture in infancy and visuospatial reasoning in children. a) While viewing speakers' faces, four-month olds spent a greater proportion-of-total-looking-time (PTLT) on a speaker's eyes, whereas 8- to 12-month olds spent greater PTLT on a speaker's mouth. In adulthood, the balance shifts back to a speaker's eyes. Reprinted with permission from Lewkowicz and Hansen-Tift (2012). b) In a developmental comparison of matrix reasoning, the authors defined an "encoding" sequence as three subsequent fixations along a row or column of the matrix problem space. Using a median split by performance, higher-performing 5-6 year-olds demonstrated approximately the same encoding prevalence as 7-8 year-olds (left). Similarly, children who received feedback on how to

complete the puzzles demonstrated more encoding behavior than those who did not (right). Reprinted with permission from Chen, Honomichl, Kennedy, and Tan (2016).

The second class of studies can only be successfully administered on older children and adults, wherein they must perform a task that explicitly requires them to make specific eye movements in accordance with task rules. Perhaps the best example is a measure of cognitive control known as the antisaccade task (Hallett, 1978; Munoz & Everling, 2004). On this task, participants are asked to fixate at the center of the screen and wait for a stimulus to flash. On prosaccade trials, they are asked to move their eyes rapidly to the target; on antisaccade trials, they must move their eyes to the mirror-opposite location of the screen. As noted previously, orienting towards a stimulus is reflexive; thus, moving one's eyes to the opposite location on the screen requires inhibitory control, as well as maintenance of the relevant task rule. Using this task, Luna and colleagues have shown that cognitive control improves gradually over childhood and adolescence, leading to fewer and fewer errors on antisaccade trials (Luna et al., 2004). In fact, this task reveals a more protracted developmental trajectory for cognitive control than most tasks involving hand movements, likely because it is particularly difficult to break the strong link between attention and gaze (Deubel & Schneider, 1996; Shepherd et al., 1986). This second class of oculomotor studies also includes investigations of the development of processing speed (Luna et al., 2004), planning (Asato, Sweeney, & Luna, 2006), language (e.g., atypical patterns in dyslexia; Tiadi, Gérard, Peyre, Bui-Quoc, & Bucci, 2016), mental imagery (Johansson, Holsanova, & Holmqvist, 2006), and spatial working memory (Luna & Velanova, 2011; see also Theewes, Belopolsky, & Olivers, 2009).

The third class of studies involves analysis of sequences of saccades around a complex stimulus. There is a smaller developmental literature using this approach than the others, but there are some examples. For example, eye movement analysis has been used to characterize differences in the way in which beginning and skilled readers approach a text. Typically, while reading, very short words are not fixated at all, while longer words almost always are, and people often go back and fixate again on words that have more letters or are more difficult to comprehend (K. Rayner, 1998). Beginning readers or less-skilled readers exhibit longer fixations, shorter saccades, and more refixations than skilled readers (Rayner, 2009). Delving more deeply into sources of individual differences in eye movement patterns during reading, several researchers have found that working memory capacity is an important factor (Calvo, 2001; Traxler, Williams, Blozis, & Morris, 2005). Thus, combining eye gaze analysis with independent cognitive measures can help us to isolate key processes underlying a complex behavior.

Researchers have also begun to use gaze measures to study the development of the ability to reason about novel problems (Chen, Honomichl, Kennedy, & Tan, 2016; French & Thibaut, 2014; Glady, Thibaut, French, & Blaye, 2012; Thibaut, French, Missault, Gérard, & Glady, 2011; J. P. Thibaut & French, 2016). We are beginning to understand the neural underpinnings of developmental improvements in reasoning ability (Dumontheil, Houlton, Christoff, & Blakemore, 2010; Wendelken, Ferrer, Whitaker, & Bunge, 2016), but these studies provide only limited insights into the underlying cognitive changes. Gaze path analyses can provide detailed insights into the strategies and approaches people use as they solve complex problems. In one recent study, Chen et al. (Chen et al., 2016) investigated the difference in reasoning strategies between younger and older children, and also the change in strategies as children received feedback throughout the session (**Figure 2b**). Using a visuospatial matrices task, they identified specific gaze patterns that they proposed reflect necessary task operations, and counted the number of times those occurred in a trial, as opposed to other eye gaze patterns. They found that older children, better-performing children, and those who received helpful feedback during task performance all demonstrated more task-specific operations than children who performed less well due to their age, skill level, or lack of feedback.

The breadth and depth of cognitive insights gleaned from gaze analyses motivates the expansion of eye-tracking methodology in several directions. Regarding gaze analyses, longitudinal studies examining the development of cognitive skills are relatively rare (exceptions include Huestegge, Radach, Corbic, & Huestegge, 2009; and Schneider, Kron, Hünnerkopf, & Krajewski, 2004). Such studies would enable us to illuminate the shift in strategies as children and adults construct new concepts, build new skills, and gain expertise across a variety of cognitive domains. Additionally, two methodologies now widely available via standard eye-tracking technology have the potential to augment the insights of gaze analyses: pupillometry and spontaneous blink rate. The analysis of pupil dilation has been used for over a century in the scientific study of cognitive processes (Kahneman & Beatty, 1966; Löwenstein, 1920; Schweitzer, 1956), but obtaining these data required hand-measurement of photographs taken of the pupil every 0.5-1 second, or the use of infrared pupillometers that obscured the participant's vision. Similarly, measures of blink rate have informed cognitive and clinical studies since the 1920s (e.g., Ponder & Kennedy, 1927), but required hand-counting of visually observed blinks, the use of electrooculography (EOG), or other custom-made devices. Now that both of these measures can be obtained with modern eye-trackers and analyzed with automated data processing software, we recommend the expansion of their use in developmental studies.

As the use of eyetrackers becomes more widespread, it is important that researchers who are just beginning to use this methodology understand both its affordances and its limitations. Just as fMRI indirectly measures brain activity by measuring blood oxygenation, necessitating that researchers mitigate and account for the effects of the physiological and idiosyncratic factors that affect blood flow, there are also many potential influences on ocular responses that must be considered (Gredebäck et al., 2009). Below, we provide an introduction to these ocular measures, the neural mechanisms they reflect, and the opportunities they present for new insights into cognition and cognitive development.

## 1 Pupil Dilation

Changes in pupil size are caused by two antagonistic muscles (**Figure 1b**): the dilator pupillae, which is located in the outer parts of the iris and dilates the pupil, and the sphincter pupillae, located in the central parts and constricting it. The constricting sphincter muscle receives input from brain systems involved in the pupillary light reflex (Loewenfeld & Lowenstein, 1993), but both pupillary muscles also receive inputs from brain systems involved in cognitive and autonomic functions (Samuels & Szabadi, 2008). As a result, changes in cognitive and autonomic activity influence pupil diameters. Pupil dilations cannot be inhibited voluntarily, although it is possible to dilate one's own pupils, for example by doing mental arithmetic (Loewenfeld & Lowenstein, 1993). Neuroscientists and cognitive psychologists have exploited the pupillary response to cognitive effort to study the unfolding of cognitive processes over time by observing fluctuations in pupil diameters. A large number of studies has used this method for at least 6 decades in human adults (e.g., Lowenstein & Loewenfeld, 1958) and – to a lesser extent – in infants (review: Hepach & Westermann, 2016). But despite its many advantages, pupillometry has been underrepresented in the study of children and adolescents so far.

Below, we first provide a detailed overview of the neural systems underlying the relationship between cognition and pupil dilation. We do so for two main reasons: First, the close relationship between task-evoked pupil dilation and its underlying neural mechanisms provides a strong argument that this method can be used with participants of any age. Second, knowledge about this relationship allows researchers to translate results about the neural system into cognitive studies, and to interpret results of cognitive studies in terms of underlying neurophysiological processes. After reviewing the neural substrates of the pupillary response, we will detail how this method has been used to study cognitive processes in adults, and how it could be used in children.

## 1.1 Pupil dilation as a Proxy of Noradrenergic Activity in the Brain

The pupils of the eye not only constrict in response to light and dilate in response to darkness; in children as well as adults, they also dilate during autonomic arousal (Beatty & Lucero-Wagoner, 2000; Granholm & Steinhauer, 2004; Samuels & Szabadi, 2008) and mental activity (Beatty, 1982; Granholm & Steinhauer, 2004; Loewenfeld & Lowenstein, 1993; Sirois & Brisson, 2014). The reason that the pupil responds to arousal and mental activity is that pupil dilation is modulated by the noradrenergic locus coeruleus (LC), as shown in **Figure 3** (Rajkowski, Kubiak, & Aston-Jones, 1993). The LC is a small nucleus in the brainstem that plays a central role in the regulation of physiological arousal (Samuels & Szabadi, 2008) and cognitive functioning (Sara, 2009). Below, we will first describe the neural structures underlying the tight relation between pupil dilation and LC activity and then turn to the role of the LC system in arousal and cognition.

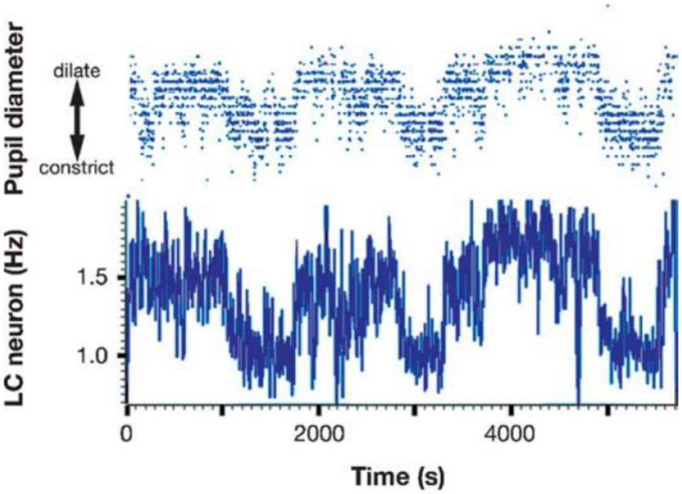


**Figure 3.** Temporal coupling between pupil diameter and firing of a single LC neuron of a monkey during performance of a signal-detection task. The relationship between LC firing and pupil diameter is mediated through the projection of the LC to the Edinger-Westphal nucleus, the origin of the pupil's parasympathetic constricting fibers, and through the influence of the LC-NE system on sympathetic nervous activity, which promotes pupil dilation. Reprinted with permission from Rajkowski et al. (Rajkowski et al., 1993).

The LC is a cluster of neurons that release norepinephrine (NE; also called noradrenaline), a neuromodulator with widespread influences on central and peripheral nervous system activity that will be described later. NE is essential for normal brain development; it is expressed in the brain even before the brain's neurons are differentiated (Herlenius & Lagercrantz, 2001). NE projections to the LC's cortical target structures are already in place at birth (Marshall, Christie, Finlayson, & Williams, 1991). Baseline levels of NE continue to change during development, increasing steadily before birth, reaching their maximum shortly after birth, and remaining largely stable throughout childhood (Herlenius & Lagercrantz, 2001; Robinson, 1975).

The relationship between the pupillary system and LC-NE activity has been established through numerous anatomical and physiological studies in both adult humans and animals. The strength of temporal coupling between LC firing frequencies and pupil diameter in monkeys is striking (Costa & Rudebeck, 2016; Joshi, Li, Kalwani, & Gold, 2017; Rajkowski et al., 1993; Varazzani, San-Galli, Gilardeau, & Bouret, 2015), and fMRI studies have shown corresponding links between LC BOLD signals and pupil diameter in humans (Alnaes et al., 2014; Murphy, O'Connell, O'Sullivan, Robertson, & Balsters, 2014). A recent study involved simultaneous recording in rhesus monkeys of pupil diameter, neuronal firing of the LC, and neuronal firing of the substantia nigra pars compacta (SNc), which is the production site of the neuromodulator dopamine (DA). This study showed that LC activity was linked to pupil dilation and the monkeys' exerted effort, providing evidence for links between LC activity and pupil dilation and between pupil dilation and cognition (Varazzani et al., 2015). LC-NE activity leads to pupil dilation because the LC has direct inhibitory projections to the parasympathetic Edinger-Westphal nucleus, where the pupil's constricting fibers originate. By inhibiting the Edinger-Westphal nucleus and the pupil's constricting muscle, LC activity therefore indirectly dilates the pupil (Beatty & Lucero-Wagoner, 2000; Loewenfeld & Lowenstein, 1993; Samuels & Szabadi, 2008). LC activity also increases activity in the sympathetic system (Samuels & Szabadi, 2008), including sympathetic fibers that innervate the pupil, resulting in additional pupil dilation (Loewenfeld & Lowenstein, 1993). A pharmacological study in humans has shown that both inhibition of the parasympathetic constricting fibers and excitation of the sympathetic dilating fibers contribute to the cognitively-evoked pupil response (Steinhauer, Siegle, Condray, & Pless, 2004).

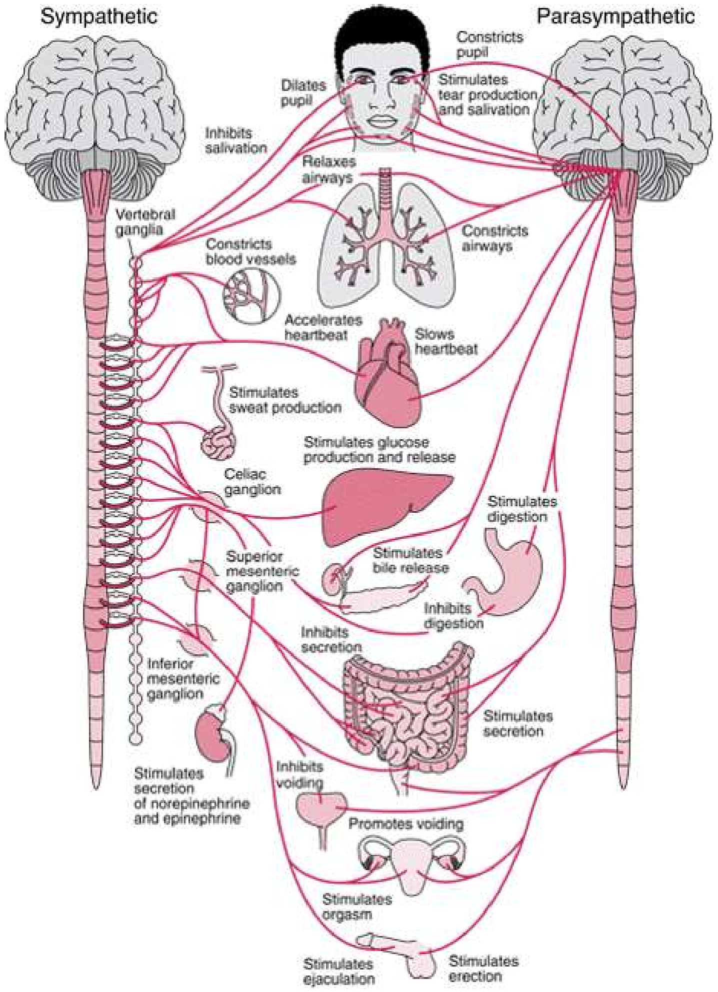


**Figure 4.** Anatomy of the autonomic nervous system and its sympathetic (fight-or-flight) and parasympathetic (rest-and-digest) branches. Post-ganglionic activity is mostly mediated by NE in the sympathetic branch and by acetylcholine in the parasympathetic branch. Many organs receive inputs from the sympathetic and parasympathetic branches, in which case functions are often reciprocal, as with pupil dilation (sympathetic dilates, parasympathetic constricts) or heart rate (sympathetic accelerates, parasympathetic slows). Reprinted with permission from the Merck Manual Professional Version, known as the Merck Manual in the US and Canada and the MSD Manual in the rest of the world, edited by Robert Porter. Copyright 2016 by Merck Sharp & Dohme Corp., a subsidiary of Merck & Co, Inc, Kenilworth, NJ. Available at http://www.merckmanuals.com/professional. Accessed April 27, 2016.

Samuels and Szabadi (2008) have provided a comprehensive review of the anatomical structures underlying the LC's role in arousal and autonomic function. NE release promotes wakefulness through the LC's dense excitatory projections to the majority of the cerebral cortex and to other structures related to alertness. At the same time, LC activity suppresses sleepiness through its substantial inhibitory projections to sleep-promoting GABAergic neurons in the basal forebrain and to other structures related to low arousal throughout the central nervous system (Samuels & Szabadi, 2008). LC activity also increases sympathetic activity and decreases parasympathetic activity via direct projections to the sympathetic and parasympathetic divisions of the spinal cord and indirect projections to various nuclei influencing the autonomic system.

As a neuromodulator of brain activity, NE influences cortical processing globally and has a crucial influence on cognitive processes. The LC is the only source of NE in the cerebral cortex (Sara, 2009), projecting to widespread but highly specific sites (**Figure 5a**). Sara (2009) and Ramos and Arnsten (2007) have provided comprehensive reviews on NE, focusing specifically on its role in attention and memory (Sara, 2009) and executive function (Ramos & Arnsten, 2007). NE's function is elucidated by its effects on sensory processing. When NE is applied to primary sensory neurons in the auditory or somatosensory cortex, spontaneous firing rates ("noise") decrease, while responses to sensory stimulation ("signal") are spared (Foote, Freedman, & Oliver, 1975; Waterhouse & Woodward, 1980). In other words, NE boosts the signal-to-noise ratio of incoming sensory information. This effect has been called "gating" because it determines which input will be processed further. Similar gating effects have been found for many target areas of the LC, including the cerebral cortex, hippocampus, midbrain, thalamus, and spinal cord (Foote, Bloom, & Aston-Jones, 1983). NE also "tunes" neuronal responses; for example, it boosts the firing rates of neurons that are selective for a line's orientation that is held in working memory, while decreasing the firing rates of neighboring neurons that represent slightly different orientations. In other words, NE narrows the tuning curve of the responding neurons, resulting in a steeper decrease in activation from the memorized line orientation compared to slightly different line orientations (Ramos & Arnsten, 2007). In frontal regions, NE's gating and tuning effects crucially influence how narrow or broad the attentional focus is. This underlies a variety of complex cognitive functions, ranging from working memory to learning, memory, reward processing, decision making, and behavioral adaptation (Sara, 2009). Tracking pupil diameter over time allows researchers to study these processes by indirectly measuring the timing of norepinephrine release in response to a challenge.

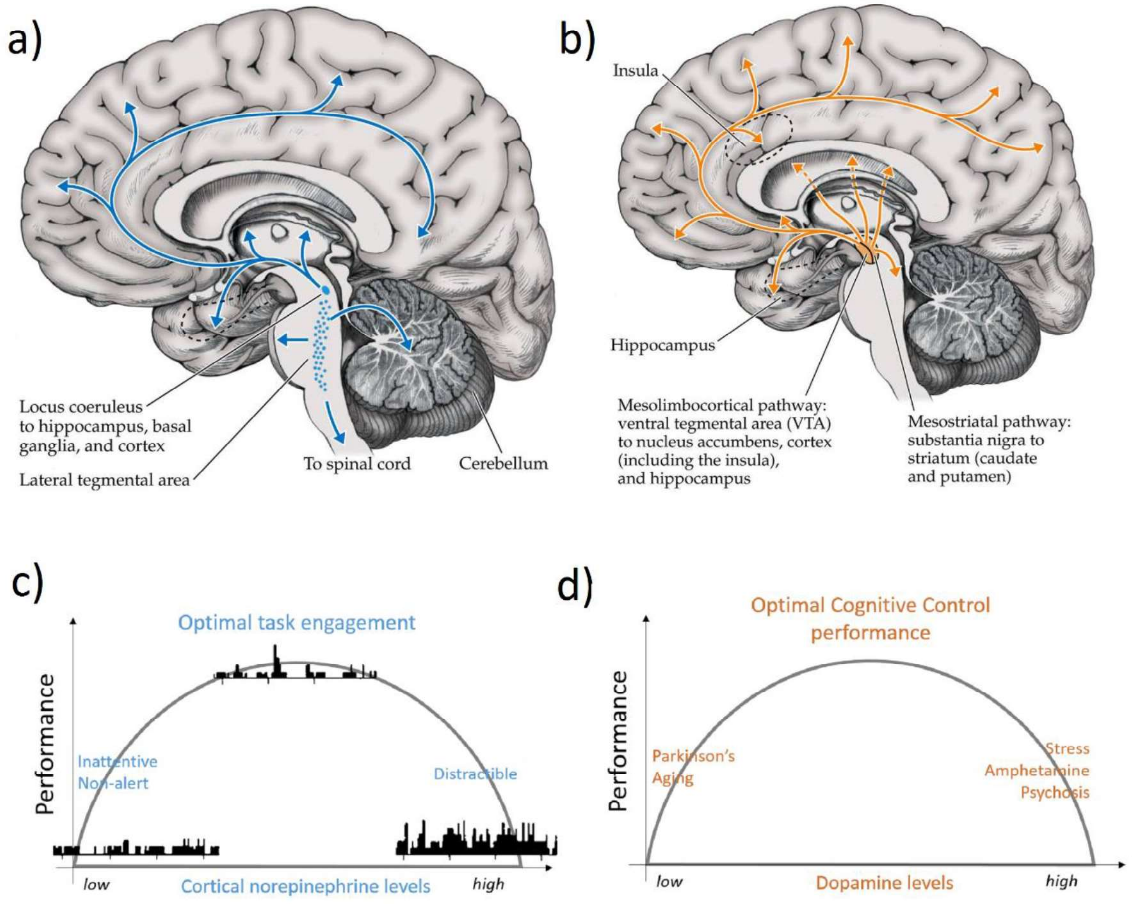


**Figure 5**. NE and DA pathways in the brain and their relationship to cognitive performance. a) The LC is the only source of cortical NE but has widespread and highly specific connections throughout the entire nervous system. The LC-NE system promotes physiological arousal and is crucial for a variety of cognitive functions, such as attention, memory, and decision making. b) DA cells in ventral tegmental area (VTA) innervate the mesocorticolimbic pathway that projects to limbic and cortical regions. In the mesostriatal pathway, the striatum receives input from DA cells of the substantia nigra. Reprinted with permission from Breedlove, Watson, and Rosenzweig (2010). c) Task performance is optimal at intermediate levels of NE, at which task-relevant stimuli elicit pronounced phasic LC responses. Low levels of NE are associated with inattentive behavior and drowsiness, and high levels with distractibility. Adapted with permission from Aston-Jones et al. (1999). d) Just as for NE, the relationship between DA levels and cognitive control performance can be described by a quadratic function. Specifically, this inverted U-shape relationship has been widely documented for D1 receptor activity and working memory performance. Adapted with permission from Goldman-Rakic et al. (2000).

The role of the LC-NE system in physiological arousal and its specific role in signal processing and cognition are intrinsically linked. This point is illustrated by a study by Aston-

Jones, Rajkowski, and Cohen (Aston-Jones et al., 1999). Nearly all of the 300+ monkey LC neurons recorded from in this study responded selectively to target stimuli in a visual discrimination task, but not to distractors, suggesting that the LC-NE system selectively boosts the processing of relevant, but not irrelevant, sensory information. As might be expected, this preferential processing of task-relevant stimuli is reduced at both extremely low and high levels of arousal. Consequently, periods of extremely reduced and extremely elevated baseline activity in the LC, resulting in drowsiness and physiological arousal, respectively, were accompanied by larger variability in the monkeys' response times, elevated false alarm rates, and largely diminished LC responses to task-relevant stimuli. Only intermediate levels of tonic LC activity, resulting in alertness and attentional focus, seem to allow for phasic LC responses to task-relevant stimuli (**Figure 5c**). Taken together, both extremely low and high levels of tonic LC activity are accompanied by a lack of task-dependent phasic LC responses as well as poor task performance, whereas intermediate levels are accompanied by phasic responses to task-relevant stimuli and good task performance. On a behavioral level, this inverse-U relationship has been famously described by Yerkes & Dodson (1908). The same relationship between LC-NE activity and task performance has since been shown in humans, using pupil dilation as a measure of LC-NE activity. Intermediate levels of LC activity, accompanied by phasic pupil dilation in response to task-relevant stimuli, were associated with better performance than both low and high tonic dilations, which were associated with diminished phasic responses (Gilzenrat, Nieuwenhuis, Jepma, & Cohen, 2010; Murphy, Robertson, Balsters, & O'Connell, 2011).

Two different theories have been proposed regarding the role of the LC-NE system in attention and behavior: the adaptive-gain theory (for a review, see Aston-Jones & Cohen, 2005) and the unexpected-uncertainty theory (Yu & Dayan, 2005). The adaptive-gain theory postulates that the LC-NE system balances the evolutionary trade-off between the exploration of unknown but potentially superior resources on the one hand, and the exploitation of well-known but potentially inferior ones on the other. According to this theory, intermediate levels of tonic LC activity, leading to focused attention and task engagement, are linked to exploitation, whereas high levels of tonic LC activity, resulting in distractibility and task disengagement, are linked to exploration (Aston-Jones & Cohen, 2005). The unexpected-uncertainty theory, on the other hand, postulates that tonic LC activity reflects uncertainty about the probability of upcoming events. According to this theory, high levels of tonic LC activity reflect high uncertainty and intermediate levels reflect a certain degree of predictability; phasic LC activity is evoked at intermediate levels, when new sensory inputs radically diverge from the participant's prevailing expectation, which is usually the case for behaviorally relevant stimuli and initiates the updating of the prior interpretation (Yu & Dayan, 2005).

### 1.1.1 Relation between pupillometry and other measures of brain activity

**Functional magnetic resonance imaging (fMRI).** Studies employing concurrent fMRI and pupillometry have shown that human pupil dilation and BOLD signals in the LC are temporally coupled, as mentioned above, supporting the use of pupil dilation as a measure of LC activity in humans (Alnaes et al., 2014; Murphy et al., 2014). Based on a review of the combined pupillometry-fMRI studies available to date, it seems that pupil dilation is also correlated with the activity of the brain regions engaged by the current task demands. We provide three examples here based on different tasks. First, on a digit-sorting fMRI task, pupil dilation was used as a regressor for the BOLD signal, and was found to be temporally correlated with activation of regions in left lateral prefrontal cortex and bilateral parietal cortex that have been implicated in working memory and cognitive control (Siegle, Steinhauer, Stenger, Konecky, & Carter, 2003). Similarly, on a gambling task, pupil dilation was found to be temporally correlated with activation of regions in bilateral posterior inferior prefrontal cortex and pre-supplementary motor area that have been implicated in decision-making (Satterthwaite et al., 2007). Third, in a multiple object-tracking paradigm, individual differences in pupil dilation between different numbers of tracked objects predicted activity in the dorsal frontoparietal attention network, including the FEF, anterior and posterior intraparietal sulcus, and superior parietal lobule, assessed in a separate session, above and beyond the number of tracked objects (Alnaes et al., 2014).

In addition to the task-related findings above, it has been shown that pupil dilation during resting-state fMRI or mental imagery is temporally correlated with activation of default-mode areas (medial prefrontal cortex, inferior parietal lobule, and junction of precuneus and posterior cingulate), and negatively correlated with activation of sensorimotor areas (Yellin, Berkovich-Ohana, & Malach, 2015). In this study, a slow buildup of activity in default-mode areas preceded pupil dilation, along with widespread BOLD suppression in the sensorimotor cortex. Given these findings, Yellin and colleagues suggested that elevated tonic LC activity, reflected in pupil dilation, is linked to the suppression of sensorimotor processes and the production of spontaneous thought. As a possible neural mechanism of this complex interplay, the authors propose a lagged connection between posterior inferior parietal lobule and the LC, acting as a neural accumulator regulating tonic LC activity and the balance between exploration and exploitation. Taken together, these studies show that the pupillary response not only correlates with LC activity, but also with any other type of cortical activity associated with ongoing thought processes.

We posit that these pupillary-brain correlations can be explained in at least two ways: (1) Pupil dilation reflects attentional focus and mental effort, and should therefore naturally correlate with the brain regions carrying out an attended task. In other words, pupil dilation and the activity of specific brain regions should be temporally coupled: as a task unfolds, arousal and mental effort, reflected in pupil dilation, wax and wane and are temporally coupled to the activity of implicated brain regions. (2) There is also a more mechanistic explanation, although the two are not mutually exclusive. NE boosts the signal-to-noise ratio in task-relevant brain regions, which is reflected in an increased task-related BOLD signal. In parallel, NE also dilates the pupils. In other words, the amplitude of pupil dilation and BOLD signal should be correlated, with higher levels of neural gain and task engagement reflected in greater pupil dilation.

Most of the studies on cognitive tasks described above used average measures of pupil dilation as regressors in their fMRI analysis, whereas the resting state study used the raw dilation time course. The correlation between the raw pupil dilation and the BOLD signal reveals an aspect of temporal coupling, but the correlation between BOLD and average pupil dilation also points to a coupling of amplitudes. When task-related pupil responses averaged over multiple trials and several seconds still show a correlation with BOLD signals, this cannot be explained in terms of temporal coupling alone. The evidence therefore seems to be consistent with both explanations.

**Electroencephalography (EEG).** As a measure of activity of the LC-NE system, pupil dilation is related not only to BOLD signals, but also to EEG signals. This section will review evidence for this relationship, with the goal of convincing readers that pupil dilation is, like EEG, a reliable measure of certain aspects of brain function. The second goal of this section is to allow readers who are already familiar with EEG to relate the two measures. A number of studies have combined pupillometry with EEG in adults, in an effort to characterize the overlaps and differences between the processes measured by the two methods. No such studies have been conducted in children so far, and more research is necessary to prove that the findings obtained with adults hold at all ages. The positive-going event-related potential (ERP) labeled the P3 has – just like task-evoked pupil dilation – been proposed as a marker of LC-NE activity (Murphy et al., 2011; Nieuwenhuis, Aston-Jones, & Cohen, 2005). The P3, one of the most studied ERPs, peaks 300-600 milliseconds after the presentation of a task-relevant stimulus and is most prominent at frontal-central midline electrodes (Sutton, Braren, Zubin, & John, 1965). The notion that the P3 reflects (phasic) LC-NE activity is supported by a number of human and animal studies (for a comprehensive review, see Nieuwenhuis et al., 2005). For example, a primate neurophysiology study found that both LC neuronal firing and the simultaneously measured P3 were selectively elicited by target stimuli in a visual "oddball" task (Aston-Jones, Chiang, & Alexinsky, 1991).

The specific relationships between the LC-NE system, P3, and pupil dilation have only recently been investigated in greater depth (Hong, Walz, & Sajda, 2014; Kamp & Donchin, 2015; Murphy et al., 2011). In one study, tonic pupil dilation prior to stimulus presentation on an auditory oddball task showed an inverse-U relationship to both the evoked P3 amplitude and to task performance, supporting the claim that P3 amplitude reflects phasic LC-NE activity and task engagement (Murphy et al., 2011), similar to what we have described for tonic pupil dilation. In addition, P3 amplitude as well as stimulus-evoked pupil dilation decreased substantially over the course of the experiment, pointing to decreasing task engagement, while pre-stimulus pupil diameters and response times increased, revealing increasing distractibility. A similar but more data-driven study showed comparable relationships between pre-stimulus pupil diameter, stimulus-evoked dilation, and a P3-like EEG component (Hong et al., 2014). This study also showed a link between stimulus-evoked pupil dilation and pre-trial EEG alpha band activity (8-12 Hertz), in that reduced alpha, indicative of elevated attention, coincided with increased stimulus-evoked pupil dilation, associated with elevated task engagement (**Figure 5c**). These results suggest a strong link between phasic pupil dilation and the P3, making pupillometry a promising alternative (or addition) to EEG studies about attention, novelty, and surprise.

However, despite the many similarities between the two measures, neither of the abovementioned studies (Hong et al., 2014; Murphy et al., 2011) found a direct correlation between them. In fact, an even more recent investigation has revealed additional discrepancies (Kamp & Donchin, 2015), suggesting that the measures reflect at least partially distinct underlying mechanisms. This is not, perhaps, surprising, given that pupillometry provides a more indirect and global measure of brain function than a specific ERP component. Taken together, both pupil dilation and P3 amplitude have been used as measures of LC-NE activity and task-directed attention and are typically elicited by similar cognitive tasks. Nevertheless, the two are not interchangeable, and further research is needed to clarify how the two measures, independently or jointly, elucidate cognitive processing.

In summary, research in human adults has started to reveal the precise relationship between pupillometry and EEG and fMRI, and demonstrates that pupillometry may augment these methods. For example, including pupil dilation as a regressor in whole-brain fMRI analysis has increased the sensitivity and specificity of the results (Siegle et al., 2003), and has revealed finer temporal patterns than would be possible with fMRI alone (Yellin et al., 2015). Additional research involving simultaneous pupillometry and brain imaging data collection is warranted to articulate more clearly the relationships between them, but the evidence to date is sufficient to conclude that pupil diameter is an indirect measure of brain function.

## 1.2 Cognitive Processes Studied with Pupillometry

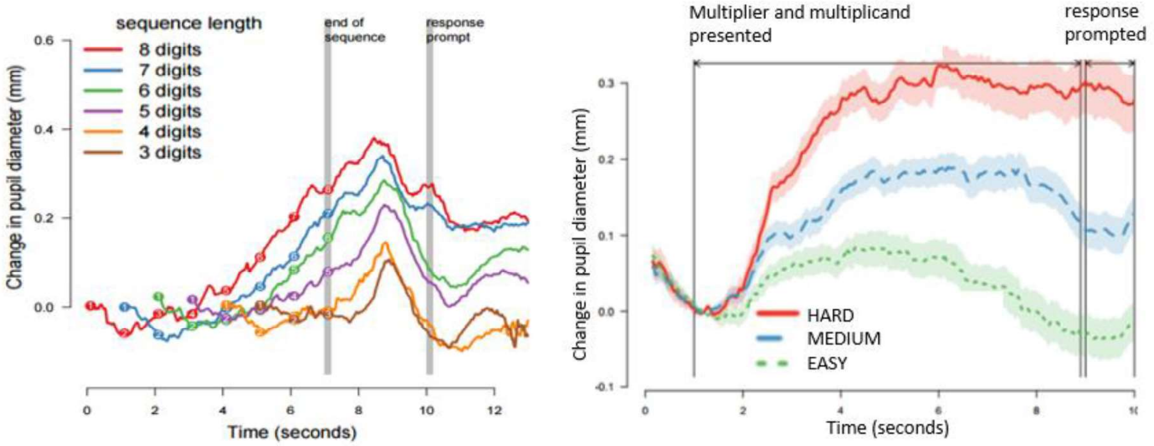
Most cognitive experiments that employ pupillometry focus on the fast, task-related LC-NE response reflected in phasic pupil dilation, while studies focusing on alertness and arousal also consider the slower, autonomic, tonic modulations reflected in baseline pupil diameters. Here, we provide only a broad overview of the topics and phenomena that have been studied by measuring phasic pupil dilation, and refer the interested reader to more detailed reviews. So far, pupillometry has been used extensively in the study of adult cognition (reviews: Andreassi, 2000; Beatty, 1982; Goldinger & Papesh, 2012; Granholm & Steinhauer, 2004; Loewenfeld & Lowenstein, 1993), and an increasing number of publications reflects its gain in popularity in infant research (Hepach & Westermann, 2016; Sirois & Brisson, 2014). To date, however, pupillometry has only been used in a handful of studies to investigate child cognition (Chatham, Frank, & Munakata, 2009; Chevalier, Martis, Curran, & Munakata, 2015; Johnson, Miller Singley, Peckham, Johnson, & Bunge, 2014; Karatekin, Marcus, & Couperus, 2007; Tharp et al., 2015).

Pupil dilation is often used as a measure of a person's mental effort when working on a task, or, on the flipside, a task's intrinsic processing demands (Beatty, 1982). A more recent computational framework has proposed that pupil dilation reflects capacity utilization, the balance between task demands and individual cognitive resources (Just, Carpenter, & Miyake, 2003). Other suggestions of how to relate the physiological pupillary response to cognitive descriptions have been proposed in terms of adaptive gain and unexpected uncertainty, as described above (Aston-Jones & Cohen, 2005; Yu & Dayan, 2003).

Although the labels for the cognitive processes underlying the pupillary response differ between theories, it is clear that pupil dilation reflects a specific, intensity- and attention-related aspect of cognitive processing. For example, a large number of studies has shown that pupil dilation scales with levels of difficulty across a wide range of tasks, including short-term memory (Kahneman & Beatty, 1966; J. Klingner, Tversky, & Hanrahan, 2011), arithmetic operations (Figure 6b; Ahern & Beatty, 1979; Hossain & Yeasin, 2014; J. Klingner et al., 2011), digit sorting (Siegle et al., 2003), sentence comprehension (Ahern, 1978), and perceptual matching (Ahern & Beatty, 1979). Across all of these tasks, more difficult problem-solving conditions elicited larger pupil dilations than easier conditions. The same has been shown in children (Boersma, Wilton, Barham, & Muir, 1970), making pupillometry a promising method with which to quantify differences in task difficulty. The relationship between the intensity of cognitive processing and pupil dilation also holds across tasks, with simpler paradigms eliciting smaller pupil dilations than more complex ones, even when the tasks stem from unrelated domains, such as when comparing a perceptual task to mental calculation and sentence comprehension (Beatty, 1982). Taken together, the fact that pupil diameter scales with task demands makes it a valuable tool for objectively measuring the intensity of cognitive processing in participants of any age.

(a) Short-term memory: Digit-span task

(b) Working memory: Multiplication task



(c) Task-relevant processing: Target detection task

(d) Cognitive control: Stroop task

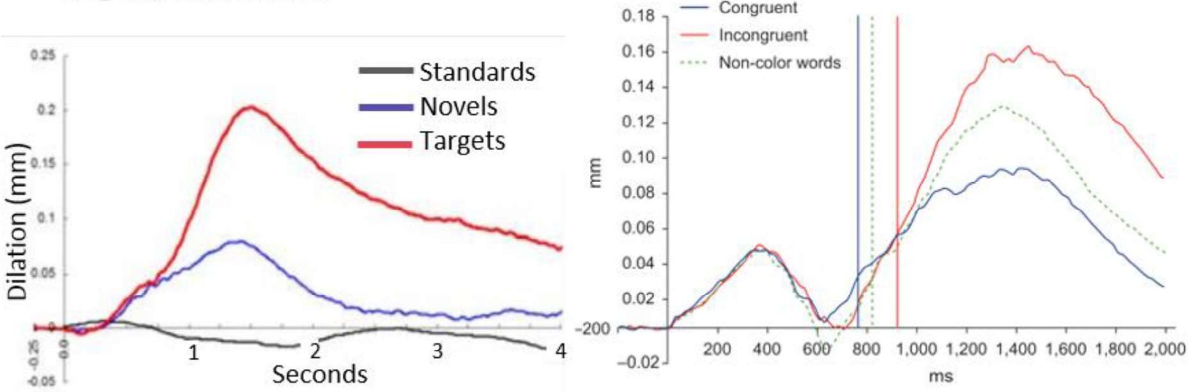


**Figure 6**. Pupil dilation scales with task difficulty in a variety of cognitive domains. a) Short-term memory: Digit-span task. Subjects saw 3-8 digits, presented sequentially for one second each, and attempted to recall all digits after a retention interval of 3 seconds. Pupil dilation increased as a function of short-term memory load. Reprinted with permission from Klingner, Tversky, and Hanrahan (J. Klingner et al., 2011). b) Working memory: Multiplication task. Subjects were asked to mentally multiply two visually presented numbers. The numbers were smallest in the “easy” condition, bigger in the “medium” condition, and biggest in the “hard” condition. Pupil dilation scaled with task difficulty and remained elevated for several seconds

after stimulus presentation. Reprinted with permission from Klingner (2010). c) Task-relevant processing: Oddball task. Subjects listened to a stream of auditory stimuli and were instructed to press a button in response to target tones only. Target tones (1500 Hz) made up 10% of the presented stimuli, 80% were standard stimuli (1000 Hz), and 10% were novel stimuli (bells, whistles, horns, etc.). There was no sign of pupil dilation in response to standard stimuli. Novel stimuli elicited a pronounced pupil dilation of more than 0.5 millimeters, but target tones elicited a much larger response of 2 millimeters, reflecting selective orientation toward task-relevant stimuli. Reprinted with permission from Book, Stevens, Pearlson, and Kiehl (2008). d) Cognitive control: Stroop task. Subjects were asked to name the color of 320 letter combinations presented for 2 seconds. In congruent trials, the colored letters formed the name of the color, whereas in incongruent trials, the letters formed the name of another color. Non-color words were used as a control condition. Pupil dilation was reduced in congruent trials relative to non-color words and was increased in incongruent trials, suggesting that pupil dilation is a sensitive measure of cognitive control. Reprinted with permission from Laeng, Ørbo, Holmlund, and Miozzo (2010).

Pupil dilation also reveals which trials in a task elicit the greatest cognitive effort, such as target tones in a target detection task (Book et al., 2008) or incongruent trials in a cognitive control task (Laeng et al., 2010; Siegle, Steinhauer, & Thase, 2004). Similarly, adults and infants show elevated pupillary responses to deviant tones embedded in a sequence of identical tones, in an auditory oddball paradigm (Wetzel, Buttelmann, Schieler, & Widmann, 2016). In children, proactive and reactive cognitive control (Chatham et al., 2009; Chevalier et al., 2015) and working memory (Karatekin et al., 2007) have been studied using pupil dilation to reveal the intensity of cognitive processing at every time point in a trial.

In addition to reflecting differences in cognitive processing between tasks, pupil dilation also reflects cognitive characteristics of the individual participant. In one study, adults with higher scores on intelligence tests showed smaller pupil dilations during a number of cognitive tasks (mental multiplication, digit span, sentence comprehension) than those with lower scores (Beatty, 1982), suggesting that more skilled participants exerted less effort to complete the task. This study highlights the relation between the pupillary response and individual differences in cognitive processing. We do not know of similar studies in children, but note that pupillary responses are a promising objective measure of cognitive effort for children who are too young to provide reliable self-assessments. Taken together with the abovementioned studies that examined the effects of manipulating task difficulty, pupil dilation seems to reflect how intensity and attention unfold over time during the cognitive process, which is shaped by task demands as well as individual differences in cognitive functioning.

Although we emphasize here studies in which pupillometry has been used to study cognition, it can also be used to study responsiveness to motivationally salient stimuli (for a review, see Sirois and Brisson, 2014). Indeed, pupil dilation has been shown to be a sensitive index of pain perception, negative emotions, sexual attraction, and subjective preferences across a range of stimuli, highlighting the point that pupil dilation reflects arousal across a range of contexts, and is not specific to cognitive challenges. Thus, tight experimental control is required to avoid potential confounds.

## 1.3 Potential of Pupillometry to Inform Developmental Research

The rich literature on task-evoked pupillometry in adult humans and in animals reviewed above lays a solid foundation for research on learning and development. While few studies have applied this technique in research with children thus far, we introduce below two lines of research that show how pupillometry has already been used to provide mechanistic insights into cognitive development. We first introduce the experimental paradigm, analyses, and results from both lines of research, and then discuss the applicability of pupillometry in developmental cognitive research.

In one line of research (Jackson & Sirois, 2009; Laeng, Sirois, & Gredeback, 2012; Sirois & Jackson, 2012), pupillometry has been used to shed light on object permanence in infants, a topic that has provoked decades of debate. In one study (Sirois & Jackson, 2012), infants' pupils were recorded while they looked at events that did or did not violate the principles of object permanence: a drawbridge that either passed through an occluded box or stopped upon touching the hidden object, similar to Baillargeon and colleagues' classic experiment (Baillargeon, Spelke, & Wasserman, 1985). In Sirois and Jackson's control conditions, the drawbridge moved in similar ways as in the experimental conditions, but no box was present (for a more detailed description, see **Figure 7a**). The authors employed functional data analysis (see following section) to precisely characterize the effects of the two experimental factors (presence of the box and rotation angle of the drawbridge) on the infants' cognitive processes at any time in the trial. The effect of the presence of the box on pupil dilation was confined to the moments at the beginning and end of the trial when the box was visible in one condition, but not in the other (**Figure 7a**, dashed line), and the effect of the rotation angle of the drawbridge was confined to moments that were similarly relevant in terms of the drawbridge's movement (**Figure 7a**, thin solid line). These findings suggest that the infants paid attention to the relevant features and events in this paradigm, and also that pupil dilation is a sensitive measure of cognitive processing even within the first 10 months of life. Next, in order to test the infants' knowledge of object permanence, Sirois and Jackson tested for an interaction between presence of the box and rotation angle of the drawbridge (**Figure 7a**, thick solid line), which would indicate that the infants processed the impossible event differently from the possible events. The fact that there was no interaction between the two factors cast doubt on the infants' knowledge that the box continued to exist after it was occluded – and of their knowledge about object permanence in general.

The second study outlined here aimed to evaluate the extent to which differences in allocation of attention at encoding could explain differences in short-term memory (STM) between children and adults. It is well-known from previous research that pupils dilate while to-be-remembered digits are presented in a digit span paradigm, plateau while the digits need to be retained, and slowly constrict as they are recalled (Cabestrero, Crespo, & Quirós, 2009; Granholm, Asarnow, Sarkin, & Dykes, 1996; Kahneman & Beatty, 1966). In the developmental study featured here, 10-year-olds and adults performed a task in which they were asked to memorize a number of digits that exceeded their STM capacity – i.e., a STM 'overload' task (Johnson et al., 2014). Children's pupils did not dilate throughout the digit presentation but plateaued around the sixth digit and constricted during the later ones (**Figure 7b**), suggesting that their attention waned prior to the end of the stimulus sequence. Adults reached a similar plateau significantly later. Notably, the position of the digit at which the dilation peak was reached predicted a child's or adult's STM capacity, measured on the standard digit span task, over and above STM capacity as measured with the eyetracking task. The authors concluded that the allocation of cognitive resources at encoding is an important factor in the development of STM (Johnson et al., 2014). With behavioral measures collected at the end of a trial, it would not have been possible to directly measure cognitive processing during encoding; by contrast, with eyetracking, it was possible to collect moment-by-moment measurements throughout the trial.

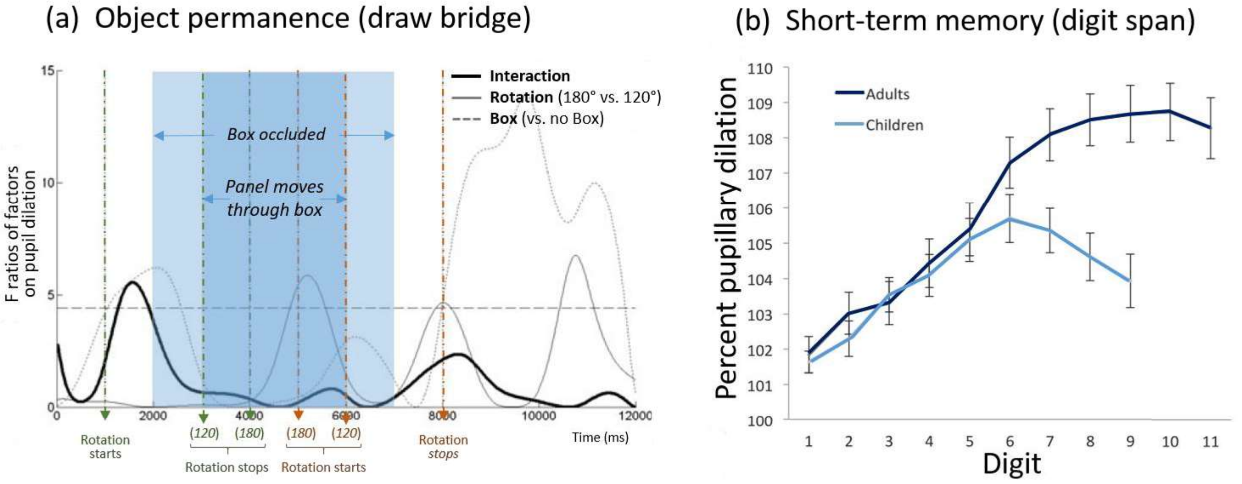


**Figure 7**. Examples for the use of pupillometry in development. a) Object permanence. Ten-month-old infants saw drawbridges that, by rotating, occluded a box behind. Infants' pupils responded to the rotation of the drawbridge, revealed by a main effect of rotation (180° or 120°), and to the presence of a box, revealed by a main effect of the presence or absence of the box, but did not respond to the violation of the principles of object permanence (box present and 180°), as would be revealed by an interaction between both. Adapted with permission from Sirois and Jackson (Sirois & Jackson, 2012). b) Short-term memory. Differences in pupil dilation between children and adults while listening to long sequences of to-be-recalled digits. The premature drop in children's pupil diameters before the end of the sequence suggests that their worse recall performance might be caused by a lack of attention during encoding and a failure to allocate sufficient cognitive resources. Reprinted with permission from Johnson et al. (Johnson et al., 2014).

Taken together, both of the studies featured here show how much detail pupil dilation can provide about cognitive processes, even when these processes have already attracted decades of scientific attention. The effects of the different trial events (movement of the bridge and occlusion of the object in the first study; presentation of each digit in the second study) could be characterized individually by measuring pupillary dilation in response to each event. The same would not have been possible with summary measures of behavior, such as looking times or behavioral responses, which reflect the final outcome of various mental processes operating at different points in a trial. Further, interpreting pupil dilations is more straightforward than looking times because of its close link to a single brain system (Aslin, 2007; Jackson & Sirois, 2009).

## 1.4 Methodology of Pupillometry

Pupil diameter is measured automatically by most conventional eyetrackers because pupil tracking is necessary to determine gaze position. Some eyetrackers are directly integrated into computer screens; others are external devices and set up below a screen or other medium of interest. Modern eyetrackers include calibration software, and pupil diameter is provided in millimeters (rather than pixels, as with older-generation eyetrackers).

Pupil diameters vary between 1.5 mm in bright light and 9 mm in total darkness. Whereas they can double or halve in size due to changes in luminance, cognitively-evoked responses are usually smaller than 0.5 mm (Beatty, 1982; Sirois & Brisson, 2014). Nevertheless, these changes can be detected reliably when luminance is held constant or when the data have been cleaned from the effects of changing luminance, for example using the method proposed by Pomplun, Sunkara, Fairley, and Xiao (2009). Some researchers conduct pupillometry experiments in dimly lit rooms with the goal of obtaining maximum amplitudes of pupil dilation. Contrary to this intuition, cognitively-evoked pupil dilation is larger in moderate light than in darkness (Steinhauer et al., 2004). (**Figures 1b and 4**). We therefore recommend conducting cognitive experiments in moderately lit rooms.

Pupil dilation can be measured at the sub-millisecond time scale, but researchers should keep in mind the response latency of the pupillary system when designing experiments. The pupil takes up to 1.5 seconds to reach maximum constriction after a sudden flash of light (Loewenfeld & Lowenstein, 1993), and tracks slow changes in luminance with a lag of about 1 second (Yellin et al., 2015). In cognitive studies, the pupil usually begins dilating immediately after stimulus presentation, but takes approximately 1-1.5 seconds to reach its maximum elicited dilation, depending on the nature and difficulty of the task (Murphy et al., 2011; also compare the different pupil dilation time courses in **Figure 6**). In order to reduce overlaps between consecutive stimuli and trials, most researchers temporally separate subsequent stimuli by about 1 second, and insert inter-trial intervals of at least 3 seconds. Nevertheless, analysis methods also exist for much faster task designs at the border of conscious perception (Wierda, van Rijn, Taatgen, & Martens, 2012).

Of relevance for developmental studies, pupil diameter changes with age. Diameter increases rapidly during the first decade of life (when measured in dim light, starting at 5.66 mm in 1-month olds), plateaus at the age of 11-15 years (with diameters between 7.10-7.45 mm in dim light), and slowly but consistently shrinks thereafter (reaching 4.5 mm in 80-year olds) (Loewenfeld & Lowenstein, 1993; MacLachlan & Howland, 2002). These changes coincide partially with changes in cerebral levels of NE (Herlenius & Lagercrantz, 2001; Robinson, 1975). For the purposes of interpreting age-related differences, relative measures of pupil dilation as compared with a baseline should therefore be used instead of absolute diameters, as described below.

Taken together, possible confounds of the task-evoked pupillary response include testing-related factors such as varying stimulus luminance, rapid presentation of stimuli, and varying lighting conditions, as well as a variety of participant-related factors, such as age, wakefulness, anxiety level, and use of pharmacological agents that affect NE levels (e.g., caffeine, marijuana, and various medications). Testing-related factors are easier to control than participant-related ones, although participants can be asked to refrain from taking stimulants or other drugs prior to testing, and a friendly testing environment can help reduce anxiety. Fortunately, many of these factors are thought to influence baseline pupil diameter rather than the phasic, task-evoked responses that are usually of interest in cognitive studies, so their effects are usually recorded but only taken into account in extreme cases. A more detailed practical guide on designing eyetracking and pupillometry experiments has been provided by Holmqvist and colleagues (Holmqvist et al., 2011).

The first step in analyzing pupillometry data is data cleaning. Dilation values reflecting measurement errors need to be identified and removed and short gaps of missing data, usually caused by blinking, can be interpolated (refer to Sirois and Brisson, 2014 for a practical guide). In the next step, absolute pupil diameters are usually transformed into relative pupil dilations. To obtain trial-wise relative pupil dilations, the pupil diameter prior to trial onset is subtracted from each data point in the trial. In the next step, trials of the same experimental conditions are averaged within subjects. EEG analysis software can be used to preprocess pupillometry data in this way, although many researchers develop their own in-house preprocessing procedures.

In the next step, data can be compared between groups or between conditions using standard statistical software. The easiest way to compare pupillary responses is to calculate average pupil dilations for a small number of pre-defined time windows (for example, stimulus presentation and response) and compare these using t-tests, ANOVA, or regression. More sensitive results can be achieved with other methods, for example functional data analysis. Here, the time course of the pupillary response is fit by a mathematical function, which is then submitted to a statistical test. The resulting test statistic is also a function over time, making it possible to determine the exact time points when a critical value is exceeded, without the need of controlling for multiple comparisons due to the number of time points. No data are lost due to averaging over large time windows, which makes functional data analysis a sensitive and precise method for analyzing pupil dilations. More detailed descriptions of this technique have been provided elsewhere (Jackson & Sirois, 2009; Ramsay, 2016; Ramsay & Silverman, 2002; Sirois & Brisson, 2014).

Many alternatives exist for analyzing pupil dilation data. As briefly mentioned above, automated deconvolution has been used to analyze a task that was at the border of conscious perception in terms of speed (Wierda et al., 2012). Principal component analysis (PCA) has been used to decompose the pupil dilation waveform and isolate dilation components associated with task performance and general cognitive abilities (Verney, Granholm, & Marshall, 2004). Time-frequency analyses, such as Fourier transform and short-time Fourier transform, have revealed specific frequency bands in the pupil dilation signal that relate to alertness (Nowak, Andrzej, & Kasprzak, 2007). The Hilbert analytic phase has been used to extract signals from the pupillary response that reflect cognitive overload (Hossain & Yeasin, 2014). A variety of methods, such as independent component analysis (ICA) (Calhoun, Liu, & Adalı, 2009), has also been proposed to facilitate the combination of pupillometry with fMRI, EEG, genetic, or other data.

As detailed above, pupil dilation provides a rich measure of neuro-cognitive processing in children and adults, and is suitable for a variety of data analysis techniques, ranging from the classic comparison of means to more data-intensive methods, such as PCA and time-frequency analysis, to cutting-edge statistical methods from the areas of machine learning and big data. However, several key points must be kept in mind with regard to study design and interpretation. First, various factors aside from cognitive demands can, if not properly controlled, influence pupil size. However, it is possible, with tightly controlled experiments, to isolate task-evoked pupillary responses that are sensitive to cognitive manipulations. Secondly, researchers must bear in mind that pupil dilation provides only an indirect measure of LC-NE activity. However, the relationship between the LC-NE system and pupil dilation has been established with a variety of methods, including human and animal studies with pharmacological manipulations and brain imaging techniques (Samuels & Szabadi, 2008; Sara, 2009), allowing researchers to draw inferences about neural processes from pupillometry data.

## 2 Spontaneous Eyeblink Rate

Dopamine (DA) is an important neurotransmitter involved in learning, working memory, and goal-oriented behavior (Westbrook & Braver, 2016). Despite decades of research on animal models and adult samples, we currently lack suitable methods for directly measuring DA activity in children and adolescents.

It has been proposed that spontaneous eyeblink rate, or the frequency at which the eyelids open and close, can serve as a non-invasive, indirect measure of DA activity in the central nervous system. In this section we review evidence that establishes the link between blink rate and the dopaminergic system, as well as studies showing the feasibility of using this marker to examine the modulatory role of DA in cognitive development and learning.

### 2.1 Spontaneous Eyeblink Rate as a Proxy of Dopaminergic Activity

Blinking serves various functions, ranging from the maintenance of ocular health to non-verbal communication. There are three main types of blinks – voluntary, reflexive, and spontaneous – which differ in their purpose and underlying mechanisms. Both reflexive and spontaneous blinks occur without volition. Reflexive blinks occur mainly as a response to environmental stimuli and cause the eyelids to shut quickly, for instance to protect the eyes from a foreign particle. Spontaneous blinks, by contrast, occur in the absence of such triggers and are characterized by a highly synchronized and transient closing and reopening of the eyelids, a movement that helps to distribute the tear film uniformly over the eye (Cruz, Garcia, Pinto, & Cechetti, 2011).

The muscles that control the opening and closing of the eyelids are the levator palpebrae superioris and orbicularis oculi muscles, respectively (**Figure 1a**). The neural processes that control the movement of these muscles during spontaneous blinking are not yet well characterized (Cruz et al., 2011; Jongkees & Colzato, 2016). Neuroimaging studies in non-human primates suggest that spontaneous blinking produces broad activation patterns across the cortex, but interestingly not in areas that respond to voluntary blink production, such as the FEF and the lateral intraparietal cortex (Guipponi, Odouard, Pinède, Wardak, & Ben Hamed, 2014). Research in a rodent model points to the spinal trigeminal complex of the medulla, which has been previously associated with the production of reflexive blinks, as a key region in the circuit that controls spontaneous blinking (Kaminer, Powers, Horn, Hui, & Evinger, 2011).

It has been hypothesized that DA modulates the frequency of spontaneous blinks indirectly, by regulating the inhibition of the spinal trigeminal complex (Kaminer et al., 2011). Although the precise neural circuitry that controls blink rate still requires further investigation, several lines of research have demonstrated a link between spontaneous blink rate and dopaminergic activity in the central nervous system. In the following two sections, we summarize key pharmacological manipulation studies and findings from populations with DA-related disorders. For a more comprehensive review on the relationship between DA activity and spontaneous blinking, we refer readers to Jongkees and Colzato (2016).

### 2.1.1 Pharmacological manipulations

Studies of pharmacological manipulations in non-human animals have provided the most compelling evidence of the relationship between blink rate and DA activity. This work demonstrates that blink rate rises steeply after the administration of DA receptor agonists such as apomorphine. Conversely, administering DA receptor antagonists leads to a notable decrease in blink rate (Karson et al., 1981). This effect is specific to DA, and not simply a byproduct of general sedation (Karson, 1983).

Subsequent research has aimed to characterize the selective roles of different DA receptors in blink rate modulation, since this information would elucidate the mechanisms of blinking and also provide vital information for using blink rate as a marker of dopaminergic function. DA is transmitted to various areas of the brain (**Figure 5b**) that have receptors of varying structural, biochemical, and functional properties. The two main families of DA receptors in the brain are D1-class receptors, abundant in prefrontal cortex, and D2-class receptors, primarily expressed in subcortical structures like the striatum (Beaulieu & Gainetdinov, 2011). The cognitive functions associated with these receptor classes will be discussed in more detail in section 2.2.

In an effort to pinpoint the receptors that underlie blink rate, researchers have examined the effects of pharmacological agents that affect D1 and/or D2 receptor function. This work has yielded mixed evidence as to which receptor class is most closely involved in modulating blink rate (**Figure 8**). Some pharmacological studies suggest D2 as the primary modulator of blink rate (Groman et al., 2014; Taylor et al., 1999). For instance, the administration of a D2 agonist, but not D1 agonist, results in increases in blink rate that correlate with PET (**Figure 8a and 8b**) and post-mortem measurements of D2-like receptor availability in the striatum of vervet monkeys (Groman et al, 2014).

In contrast, other studies point to D1 as the primary modulator of blink rate (Kotani et al., 2016; van der Post, de Waal, de Kam, Cohen, & van Gerven, 2004). The most compelling evidence comes from a study where the systematic administration of a D1-agonist produced dose-dependent increases in blink rate in marmosets (Kotani et al., 2016). A similar dose-dependent blink rate response (**Figure 8c**) was observed after the administration of a D1/D2 non-selective agonist that has been used on previous seminal blink rate studies (e.g. Karson, 1983). Importantly, the effect on blink rate of this D1/D2 agonist was selectively reversed only with the administration of a D1-antagonist (**Figure 8d**). Lastly, other studies posit an interaction or common effect of D1 and D2, in addition to independent modulatory effects of each receptor (Elsworth et al., 1991; Karson, 1983; Kleven & Koek, 1996). These conflicting results underscore the complex relationship between blink rate and DA activity (for a review, see Jongkees & Colzato., 2016). Much of the evidence suggests that both D1 and D2 receptors can modulate spontaneous blinking, and Jongkees and Colzato propose that the effects of D1 receptors on spontaneous blink rate might only be produced at certain dosages of pharmacological manipulations, whereas tonic blink rate might be more closely related with D2 activity in the basal ganglia more generally. However, this account may not be consistent with the developmental trajectories of DA tone, expression of D2 receptors, and spontaneous blink rate: DA tone and expression of D2 receptors in the basal ganglia peak during adolescence (Ernst et al., 2009), whereas blink rate, as far as we can tell from cross-sectional samples, increases from infancy to adulthood (Zametkin et al., 1979). Given these conflicting accounts about D1 and D2 receptor involvement, we can conclude only at this point that spontaneous blink rate is modulated by central dopaminergic activity.

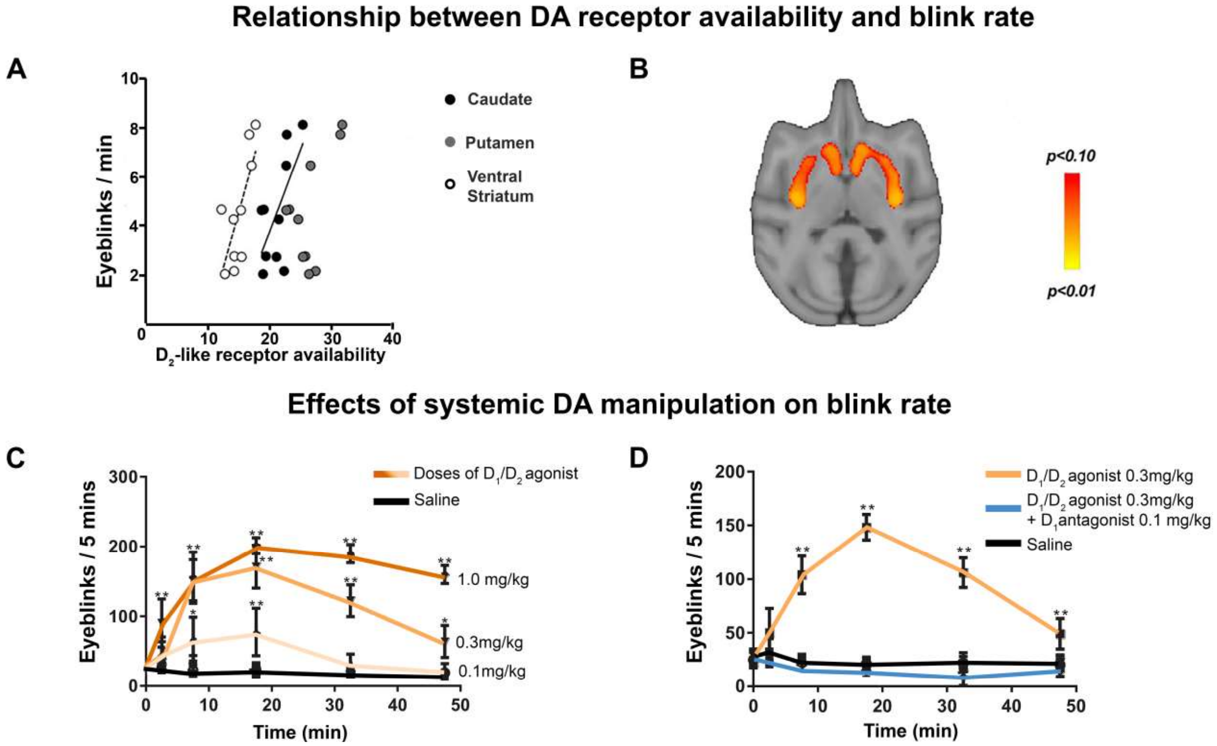


**Figure 8.** Relationship between blink rate and DA receptor activity. a) Blink rate is positively related with PET measures of $D_2$-like receptor availability in the ventral striatum (white circles) and caudate nucleus (black circles), but not putamen (gray circles). These relationships were not observed with $D_1$-like receptors (not shown). b) Statistical map (p-values) of the voxelwise linear regression of blink rate on $D_2$-like receptor availability from (a) overlaid on the striatal volume of the vervet monkey's MRI template. Adapted with permission from Groman et al. (Groman et al., 2014). c) Systemic administration of apomorphine, a non-selective DA agonist, increased blink rate in a dose-dependent manner (orange lines) above baseline levels (saline administration, black line) in marmosets. d) This effect was only reversed with the administration of SCH39166, a $D_1$-antagonist (blue line), but not with the administration of

haloperidol, a $D_2$-antagonist (not shown). Adapted with permission from Kotani et al. (Kotani et al., 2016).

### 2.1.2 Clinical studies

Patients with conditions that are known to compromise the dopaminergic system show irregularities in blink rate. For instance, low blink rate is seen in patients with Parkinson's disease (PD), a disorder that results from the loss of dopaminergic cells in the substantia nigra (Agostino et al., 2008). Medications that raise DA levels also increase blink rate in these patients (Bologna, Fasano, Modugno, Fabbrini, & Berardelli, 2012). By contrast, schizophrenia and Tourette's syndrome – conditions linked to elevated dopaminergic activity (Chan & Chen, 2004) – are associated with elevated blink rate (Tharp et al., 2015). The severity of the symptoms in these conditions is positively correlated with blink rate, a relationship that is also present in pediatric patients with early-onset schizophrenia (Caplan & Guthrie, 1994).

Although the precise influence of each dopamine receptor on the neural circuitry that controls spontaneous blink rate remains an open area of research, findings from studies with pharmacological manipulations and clinical samples with compromised dopaminergic function support the viability of blink rate as a biomarker of central dopaminergic activity. The fact that we can acquire blink rate data non-invasively makes it possible to assess DA-related cognitive functions in adulthood and development.

## 2.2 Cognitive Processes Studied with Blink Rate

Blink rate has been used as a peripheral measure of DA involvement in various cognitive and affective functions (Colzato, Van Den Wildenberg, Van Wouwe, Pannebakker, & Hommel, 2009; den Daas, Häfner, & de Wit, 2013; Fukuda, 2001; Karson et al., 1981; Oh, Han, Peterson, & Jeong, 2012; Smilek, Carriere, & Cheyne, 2010; Tharp & Pickering, 2011; Wiseman & Nakano, 2016). Cognitive control and learning have been heavily represented in the literature, and thus will be the focus of this section.

Our ability to control impulses, maintain long-term goals, and flexibly adapt to changing rules from the environment are all important aspects of cognitive control. DA is an important neuromodulator of fronto-striatal circuits that support these functions. It has been proposed that during goal-oriented behavior, DA aids in the maintenance of abstract goals in higher levels of the cognitive control hierarchy, while also allowing flexibility in updating lower-level rules guiding attainment of subgoals (Puig, Rose, Schmidt, & Freund, 2014; Westbrook & Braver, 2016). Just as for NE (**Figure 5c**, see pupillometry section), the relationship between cognitive control and DA is often described with an inverted U-shape function (Cools & D'Esposito, 2011), as shown in **Figure 5d**. There is some evidence that this relationship can be captured with blink rate, for example in the context of cognitive flexibility on a divergent thinking task (Chermahini & Hommel, 2010), as well as performance on an attentional task subsequent to cognitive depletion (Dang, Xiao, Liu, Jiang, & Mao, 2016). Most of the research in this domain has associated baseline blink rate with cognitive control, often showing a positive linear relationship.

Different functional roles related to cognitive control have been ascribed to the two DA receptor families. For instance, D1 signaling is traditionally associated with maintenance of representations of long-term goals, while D2 signaling in the striatum is associated with an increase in flexible processing and distractibility (Puig et al., 2014; Takahashi, 2013; Westbrook & Braver, 2016). In a recent study, Zhang and colleagues (Zhang et al., 2015) tested the relationship between baseline blink rate and these dimensions of cognitive control in a sample of young adults. They found that higher blink rates predicted better performance on set-shifting and Stroop tasks, but worse performance on an updating task that taxes working memory (3-back). These and other findings indicate that higher baseline blink rate is associated with better cognitive flexibility but worse maintenance (Dreisbach et al., 2005; Müller et al., 2007; Tharp & Pickering, 2011). However, other research has shown that higher blink rate at baseline is related with lower distractibility on tasks that place high demands on working memory (Colzato, Slagter, Spapé, & Hommel, 2008). Research isolating cognitive flexibility, updating, and maintenance processes is needed to better understand which aspects of the modulatory effect of DA on cognitive control can be captured with blink rate, which will be essential for examining how DA influences the developmental trajectories of these different processes.

In addition to resting blink rate, some paradigms lend themselves to measure task-evoked blink rate, which could act as an indirect measure of phasic DA and could provide additional insights into the relationship between DA and cognitive control. For example, phasic DA release is associated with "Go" signals that increase the selective updating of contextual features that are relevant for ongoing goal-directed behavior (Frank, Loughry, & O'Reilly, 2001; Westbrook & Braver, 2016). A similar phenomenon has been captured with task-related blink rate on a Flanker task, wherein the occurrence of spontaneous blinks on one trial predicted the exertion of greater control on the subsequent trial (van Bochove, Van der Haegen, Notebaert, & Verguts, 2013). These findings suggest that task-evoked blink rate can capture functions typically associated with phasic DA release, and point to the feasibility of using similar paradigms to examine, on a moment-by-moment basis, how differences in the functioning of the DA system contribute to age-related differences in cognitive control – in particular, the ability to proactively update the level of control needed given changing task demands.

Beyond cognitive control, DA also supports learning. Striatal DA activity has been linked to reinforcement learning, as it signals reward prediction errors, exhibiting a large phasic DA release when rewards are greater than expected, and firing rates below baseline when rewards are smaller than expected (Bayer & Glimcher, 2005; Glimcher, 2011; W. Schultz, Dayan, & Montague, 1997). Additionally, lower levels of tonic DA have been related to better ability to learn from negative outcomes (Pessiglione, Seymour, Flandin, Dolan, & Frith, 2006; van der Schaaf et al., 2014). This DA-related effect has been studied using blink rate. Slagter and colleagues (Slagter, Georgopoulou, & Frank, 2015) found that, compared to participants with a higher blink rate at rest, participants with a lower one learned more from negative than positive outcomes in a probabilistic learning task. In another study, positive feedback was found to mediate the relationship between PET measures of D2-like receptor availability and performance on a reversal-learning task (Groman et al., 2014). Blink rate mediated the relationship between the availability of D2-like receptors and ex-vivo measures of D2-like receptor density, thus providing a crucial link between these measures of DA and behavioral measures of reward sensitivity. Taken together, these studies show that blink rate can be used as a measure of DA-related learning, and could help bridge neuroimaging and direct measurement of DA activity.

One view of DA's role in learning is that it codes the consequences of prediction and learning signals, and not the prediction errors themselves. Instead, bursts of phasic DA release signal the salience of an incentive and enhance the 'wanting' of a reward (Berridge, 2007). This hypothesis has not been assessed with blink rate in the context of learning, but Peckham and Johnson (2015) showed the feasibility of using event-related blink rate as an indirect measure of phasic DA, in the context of reward processing in bipolar patients and healthy controls. In both groups, receiving a reward increased blink rate above resting levels, but to a lesser extent in anticipation of having to exert effort to obtain a reward (Peckham & Johnson, 2015). This study shows an exciting possibility of using task-evoked blink rate as a proxy for phasic DA release during reward processing, which could inform current hypotheses about the functional role of DA in learning. In addition, this study shows the possibility of examining mechanisms underlying developmental differences in decision-making, for example during adolescence.

In summary, there is converging evidence that blink rate can be a sensitive measure of DA involvement in certain aspects of cognitive control (e.g., cognitive flexibility) and learning, producing findings that are consistent with PET and research with laboratory animals. Designing paradigms that would allow us to measure both resting and task-evoked blink rate could help to further elucidate whether blink rate can be used to measure the effects of phasic DA release on DA tone, and to better understand individual differences and age-related changes in goal-directed behavior.

## 2.3 Potential of Blink Rate to Inform Developmental Research

The developmental trajectory of blink rate from infancy to adulthood has been primarily characterized with cross-sectional samples (for reviews, see Cruz et al., 2011; and Jongkees & Corzato, 2016). Based on the findings to date, even a fetus is capable of producing spontaneous blinks starting during the third trimester of pregnancy. Neonates and infants have a very low blink rate, blinking on average less than 3 times per minute. Blink rate increases during childhood (6-8 blinks/min) (Lavezzo, Schellini, Padovani, & Hirai, 2007; Zametkin, Stevens, & Pittman, 1979), and stabilizes at adult levels by late adolescence (10-20 blinks/minute).

The fact that blink rate can be obtained at different ages has enabled cross-sectional comparisons and the (non-invasive) study of the role of DA in typical and atypical cognitive development. The following studies represent areas of research that have used blink rate in this manner, and that can be most closely related to the adult literature; these include cognitive control, decision-making, and learning.

As highlighted in the previous section, DA is strongly implicated in cognitive control, which undergoes important changes in childhood and adolescence (Bunge & Wright, 2007; Casey, Galván, & Somerville, 2015; Luna, Marek, Larsen, Tervo-Clemmens, & Chahal, 2015). The development of the prefrontal-striatal networks in part drives age-related improvements in these abilities (Casey et al., 2015). Dynamic changes occur in the DA system in childhood and adolescence that may contribute to the development of cognitive control. For example, an increase in striatal DA activity during adolescence (Padmanabhan & Luna, 2014) could be related in part to enhanced flexibility in learning that has been observed during adolescence (Johnson & Wilbrecht, 2011).

Although no studies to date have used blink rate to measure DA and the development of cognitive control longitudinally, Tharp and colleagues (Tharp et al., 2015) provided evidence that blink rate can be used as a measure of DA function in childhood. They used blink rate both at rest and during performance of a task that required rule switching to differentiate between typically developing children and children with Tourette syndrome (TS), a condition associated with elevated levels of DA (Albin & Mink, 2006; Singer, Butler, Tune, Seifert, & Coyle, 1982). Compared to controls, TS patients in Tharp and colleagues' study were more likely to apply incorrect rules when required to switch flexibly between two rules from trial to trial. Patients also exhibited uniformly high blink rates, wherein they failed to show the pattern that typically developing children showed, of a gradation of increase in blink rate as a function of task difficulty (**Figure 9a**). This effect was observed even in unmedicated children, and was selective to blink rate since no associations were found with the amplitude or timing of pupil dilation. This study provides evidence that blink rate can be used to examine individual differences in DA functioning and engagement of cognitive control in various pediatric populations.

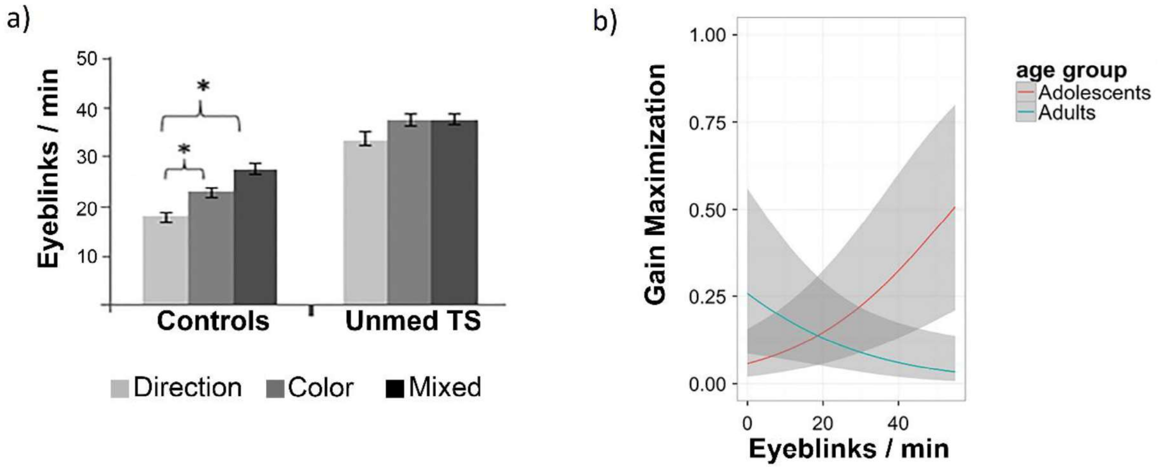


**Figure 9**. Measurement of spontaneous eyeblink rate in developmental studies. a) Blink rate during performance of a task that requires rule switching flexibility can be used to differentiate between typically developing children and children with Tourette syndrome (TS), a condition associated with elevated levels of DA. Patients showed higher average blink rate at rest as did the control subjects, and also did not show task difficulty-related increases in blink rate. Adapted with permission from Tharp et al. (Tharp et al., 2015). b) Relationship between blink rate at rest and age-related differences in reward-seeking behavior on a risky decision-making task. Higher blink rate predicted increased use of a gain-maximizing strategy for adolescents but not for adults. Adapted with permission from Barkley-Levenson and Galván (Barkley-Levenson & Galván, 2016).

Research on rodents and non-human primates has shown that cortico-striatal circuits that are influenced by DA undergo dramatic, non-linear changes during development. Adolescence is a period of heightened plasticity of the DA system, with a marked proliferation of D2 receptors in the prefrontal cortex and striatum (Ernst, Romeo, & Andersen, 2009). It has been proposed that the peak in striatal DA activity during puberty influences reward processing and sensation seeking (Padmanabhan & Luna, 2014). Barkley-Levenson and Galván (Barkley-Levenson & Galván, 2016) provided the first evidence that blink rate could be a powerful measure of age-related differences in reward sensitivity between adolescents and adults. Specifically, they used a risky decision-making task to test whether blink rate would predict how often a participant would choose options that would maximize their gains, irrespective of the probability of winning. Even though baseline levels of blink rate did not differ between adolescents and adults, higher blink rate only predicted more use of gain-maximizing strategies in the adolescent group. Considering that there was no group difference in baseline blink rate, it would be intriguing to test in future studies whether task-related blink rate could capture DA contributions, as well as age-related differences in the ability to update decision-making strategies based on past gains or losses.

The type of learning that could guide shifts in decision-making would in part require building associations between cues, actions, and outcomes over time – in other words, learning rules. Werchan and colleagues (Werchan, Collins, Frank, & Amso, 2015) have provided one of the first demonstrations that blink rate could be a valuable tool to test DA involvement in rule learning. They found that eight-month-olds were capable of learning implicit hierarchical rules, and that trials with higher-order rule switches elicited higher blink rates. Importantly, this relationship with blink rate was only present in the second half of the experiment, when the rule sets had been learned. This learning process relies on DA-innervated pathways in adults and other species (Money & Stanwood, 2013; Rothmond, Weickert, & Webster, 2012; Weickert et al., 2007). Even though the blink rate results of this study suggest that the DA system has a functional role in infancy similar to its role in adults, the functional properties or development of the DA system in the first years of human life is not well understood (Money & Stanwood, 2013; Rothmond et al., 2012; Weickert et al., 2007). There is indirect evidence from longitudinal research that there might be rapid changes occurring in this system, given the observation of an increase in resting blink rate between four and 12 months of age (Bacher, 2014).These findings are encouraging in that they suggest that blink rate can be used as a marker of DA across development, beginning in infancy.

In conclusion, these studies show the feasibility of using blink rate to indirectly measure the involvement of DA in cognitive development in both typically developing children and pediatric patient populations. Given the widespread contributions of DA to cognition, there are several lines of research that could benefit from the measurement of blink rate. An extension of the prior literature cited here could be to characterize the involvement of DA in the development of cognitive flexibility, which has been found to follow a non-linear developmental trajectory, with adolescents outperforming adults (Johnson & Wilbrecht, 2011; Simon, Gregory, Wood, & Moghaddam, 2013).

Blink rate measurement could also inform the study of the development of decision-making and learning. As shown by Barkley-Levenson and Galvan (Barkley-Levenson & Galván, 2016), blink rate is associated with age-related differences in reward-seeking behavior. Future research could evaluate whether task-evoked blink rate in decision-making paradigms has the temporal resolution needed to capture phasic changes in DA release, which are known to occur during reward processing (Wolfram Schultz, 2013). If so, this metric could prove useful for assessing changes in decision-making flexibility (i.e., change in decision-making strategies) as well as for characterizing reinforcement learning (e.g. how past outcomes affect future strategy use) during different points in development. In addition, blink rate could be a suitable method to examine individual differences in motivation during learning. It has been hypothesized that DA translates incentives into motivation, which can direct working memory allocation (Westbrook & Braver, 2016). Blink rate could also be a useful method to assess whether incentive-related changes in DA (as measured by phasic blink rate) during academic learning are predictive of future learning success. If so, blink rate could be a complementary ocular measure to improve currently available gaze-based computer tutors (e.g., Gütl et al., 2005), and a valuable method for education-based research.

Although blink rate is a promising measure for studying the development of DA-related cognitive processes, there is a strong need for a longitudinal study to assess the developmental trajectories of blink rate from childhood to adulthood. To date, only a handful of studies have measured blink rate longitudinally, and these have involved time points that were rather close in time (e.g., Bacher, 2014). Longitudinal studies would not only enrich our understanding of typical development, but could also be useful to measure outcomes of psychiatric treatment in pediatric populations, and also potentially facilitate earlier detection of risk factors for later substance abuse problems, such as enhanced sensitivity to cues of rewards and blunted D2 activity (Berridge, 2007; Goldstein & Volkow, 2011; Yin & Knowlton, 2006).

## 2.4 Methodology of Spontaneous Blink Rate

The studies reviewed thus far suggest that blink rate, as measured during paradigms known to be sensitive to DA modulation, can be a promising tool to capture DA function throughout development and adulthood. Below, we describe how spontaneous blink rate can be estimated with an eyetracker.

Some of the most popular methods to record blink rate are infrared eyetrackers, EOG, electromyography (EMG), EEG, and video cameras. Besides video cameras built into a computer, eyetrackers are arguably the most cost effective and convenient method to use with special samples. Since blinks are often characterized as a gap in data recording, traditional eyetrackers have not enabled researchers to differentiate with good fidelity whether loss of data is due to blinks, muscular artifacts (e.g. as caused by yawning) or head movements away from the visibility of the eyetracker lens. However, newer models have built-in technology to account for these issues, such as automatic adjustments for small head movements, concurrent measurement of pupil dilation, and video cameras that can be used to perform manual inspections of eyelid closures in a subset of the data (e.g., Pedrotti, Lei, Dzaack, & Rötting, 2011). Even though combining these features can improve blink detection accuracy, it is still possible to misclassify blinks that are reflexive or voluntary as spontaneous blinks. The development of detection algorithms with higher precision, as those employed with other ocular recording methods (e.g., Gehricke, Ornitz, & Siddarth, 2002; Królak & Strumiłło, 2012), is needed to increase the validity of the calculation of blink rate from data that can be obtained from eyetrackers.

Eyeblinks are defined in eyetracking data as continuous periods of approximately 100-500 msec (corresponding to the typical range of spontaneous eyeblink durations) during which the coordinates or pupil diameters of the eyes are not recorded. This threshold could vary depending on the sampling rate of the apparatus or other methodological considerations (Jiang, Tien, Huang, Zheng, & Atkins, 2013; Siegle, Ichikawa, & Steinhauer, 2008). Blink rate is then calculated as the number of blinks occurring over a specified time interval and is usually expressed in terms of number of blinks per minute (Holmqvist et al., 2011). An inverse measure, which is used less often, is the interblink interval, which refers to the time between blinks.

Another important consideration is when to measure blink rate. The studies we have reviewed have measured blink rate either at baseline or during performance of a task. Baseline blink rate is used as a proxy of tonic DA levels, whereas blink rate recorded during performance of a task is considered to reflect phasic DA. A typical paradigm for acquiring baseline or tonic measures involves having participants look at a fixation cross on the center of a dark computer screen for 5 to 6 minutes (Holmqvist et al., 2011) and as little as 3 minutes (Zaman & Doughty, 1997) to be able to capture fluctuations in blink rate that can naturally occur. Phasic or task-related blink rate is captured in response to a task manipulation, either in an event-related or block design, and the data are averaged over static or moving windows of varying lengths, and as little as 30 seconds (Peckham & Johnson, 2015; Siegle et al., 2008; Tharp et al., 2015; Werchan et al., 2015). Alternatively, some studies have also treated blink rate as time series data, which facilitates examining event-related blink rate modulations during natural viewing (e.g., Shultz, Klin, & Jones, 2011).

Several methodological factors need to be taken into consideration in order to maximize the interpretability of spontaneous blink rate measures. These include controlling for environmental and subject-specific variables that could affect blinking behaviors. Specifically, factors that affect the moisture level of the eye, such as dry eye conditions, seasonal allergies, temperature and humidity of the room, and use of contact lenses directly affect how much a person blinks to restore optimal levels of eye moisture (Al-Abdulmunem & Briggs, 1999; Cruz et al., 2011; Doughty, 2001). In addition, the presence of air pollutants, like cigarette smoke, elicit reflexive eyeblinks (Holmqvist et al., 2011). These factors can be controlled by maintaining the room's temperature and air quality stable, asking participants to wear glasses instead of contact lenses, and if possible rescheduling sessions for participants who are suffering from seasonal allergies or colds. Blink rate can also be affected by sleep deprivation (Barbato et al., 2000), as well as certain medications and stimulants that affect the DA system, like caffeine, cannabis, and antipsychotics (Holmqvist et al., 2011). It is recommended to ask participants to get a good rest before a session and to abstain from use of these drugs to minimize the acute effect of usage. In situations when it is not feasible for the participant to withhold medical treatment, the inclusion of unmedicated patients or a dose-response analysis would be needed (e.g., as done in Tharp et al., 2015).

Blink rate is also susceptible to confounds that can arise from experimental design. For example, blinking frequency increases as a factor of time in long experiments because it can affect the moisture of the eyes and also produce fatigue, both factors that modulate blink rate (Stern, Boyer, & Schroeder, 1994). Frequent breaks can eliminate these problems (Holmqvist et al., 2011). Also, visual designs that prompt large saccades or involve complex visual processing could increase blink rate irrespective of the mental workload (Al-Abdulmunem et al., 1999; Cruz et al., 2011). Thus, if measuring event-related blink rate, we recommended presenting visual stimuli centrally on the screen to minimize the number of saccades. Finally, the time of day when the session takes place can also affect results, since blink rate increases in the evening (Barbato et al., 2000). It is typically recommended that sessions take place before 5 pm.

Even when controlling for environmental and experiment design factors, a special population for whom blink rate is still challenging to obtain are infants and young children. Changes in behavioral states could alter the rate of blinking and data validity could also be affected by artifacts or gaps in the data caused by excessive motion (see L. Bacher & Smotherman, 2004 for a review of blink rate in infants). Recent technical advances such as automated body-movement adjustments and lighter, head-mounted gear can alleviate certain challenges associated with head movements (e.g., Smith, Yu, Yoshida, & Fausey, 2015), but the temporal requirements of maintaining fixed gaze for a tonic measurement may still render the data unreliable. It is therefore recommended to also carefully monitor the behavioral state of the infant or child during ocular data acquisition. Despite these cautionary notes, blink rate is still a measure that could further our understanding of the development of the dopaminergic system, as well as DA contributions to goal-oriented behavior and learning during development.

## 3 Conclusion

We have presented evidence that pupillometry and blink rate are, like eye gaze, sensitive measures of specific aspects of cognitive processing. All three ocular measures can be collected in a single session and with the same equipment, and can provide complementary information about cognitive processing. Although pupillometry and blink rate have been used extensively in adults, and to a lesser extent in infants, they have been used much less in children and adolescents – and even then, only rarely for the study of cognition. These gaps in the literature provide an opportunity to augment behavioral and brain imaging methods in studying how the mind changes over a broad developmental window.

### 3.1 Complementarity of Ocular Measures

Broadly, eye gaze can provide a moment-to-moment measure of the focus of attention, and can reveal which cognitive strategies are employed in complex tasks. Pupil dilation provides a window into the brain's LC-NE system, and reveals the subjective difficulty of cognitive tasks and the intensity of current cognitive processing. Blink rate is modulated by DA, and has been used to study cognitive control, learning, working memory, and decision making. The differences and commonalities between these measures are described further below.

Eye gaze metrics are especially suited to revealing aspects about attentional focus during visual processing. Saccades and fixations can reveal what parts of the displayed information are most salient, and to what extent and in what order information is processed within a complex stimulus set. Eye gaze can also reveal how several pieces of information are integrated or compared during a task. Unlike eye gaze measures that are well suited for visually complex tasks, pupil dilation and blink rate are best captured in tasks with auditory stimuli or simple visual design that minimize sources of artifacts, such as eye movements and large luminance contrast.

Pupil dilation and blink rate can be recorded simultaneously, given that the same task designs can be optimized for both; thus, it is possible to ask how the two measures are related. DA and NE are closely related catecholamines that modulate neural activity in overlapping brain regions (**Figures 5a and 5b**) and are associated with overlapping functional roles in cognition (Arnsten & Li, 2004; Meindertsma, 2014; Sara, 2009). Given the strong relationship between these neuromodulators, could we expect to tease apart DA and NE effects with blink rate and pupil dilation? Various research suggests that pupil dilation and blink rate do in fact reflect different processes that are uniquely associated with DA and NE (Siegle et al., 2008; Tharp et al., 2015; van Bochove et al., 2013). For example, phasic changes in pupil dilation have been associated with sustained processing of information, often associated with NE (e.g., Siegle et al., 2008), whereas phasic blink rate may respond to changing needs for flexible updating of representations, associated with DA (e.g., Tharp et al., 2015). In a working memory task, during which pupil dilation and blink rate were recorded concurrently (Siegle et al., 2008), pupil dilation but not blink rate scaled with the working-memory load of the task, as has been observed previously (Johnson et al., 2014; Kahneman & Beatty, 1966), whereas the timing of blinks but not pupil dilation was related to the presentation of new information, as has also been observed previously (e.g., Tharp et al., 2015).

Although the studies described above show that blink rate and pupil dilation are not redundant measures, there are likely cases in which they parallel one another closely, given that DA and NE release occur in tandem in many circumstances. Indeed, tonic levels of both catecholamines vary as a function of arousal, with low levels during drowsiness, intermediate levels during ‘alert interest’, and high levels during extreme stress (Robbins and Arnsten, 2009; see also **Figures 5c and 5d**). Although NE is released in response to task-relevant or surprising stimuli (Aston-Jones & Cohen, 2005; A. Yu & Dayan, 2003), whereas DA is released in response to reward-predicting stimuli or unexpected rewards (Glimcher, 2011), these conditions co-occur in many tasks. For example, a pupillary response is observed when participants receive an unexpected reward (Einhäuser, Koch, & Carter, 2010) – the classic scenario for DA release. However, the pupillary response in these situations can also be explained as the response to a surprising event. Thus, although pupil diameter and blink rate have different biological underpinnings and can be dissociated in many circumstances, this is not to be expected under conditions that elicit both NE and DA release.

Studies employing both pupillometry and measurement of eye blink rate should be conducted in developmental samples, with the goal of shedding light on the developmental trajectories of the NE and DA systems. It is conceivable that the two neurotransmitter systems mature at different rates (see Sections 1.1, 2.1, and 2.3) and that the balance between them changes during development. Longitudinal or cross-sectional studies with surprising as well as reward-predicting elements might be able to shed light on such developmental trajectories through the measurement of pupil dilation and blink rates. In addition, such studies would be crucial to determine the extent of overlap between the NE and DA systems during development.

Although eye gaze, blink rate, and pupil dilation seem to be able to provide complementary information about cognitive processing, only a small number of studies has used more than one measure simultaneously so far (e.g., Siegle et al, 2008). Employing multiple eye-tracking measures in the same task would provide the unique opportunity to assess mutual relationships between them and also to directly compare the assessments of similar cognitive processes. This is especially important because all three measures have independently been used as measures of attention, employed strategies, and effortful cognitive processing. More studies are needed to identify circumstances under which these measures may reveal either complementary or redundant information about cognitive development.

## 3.2 Next Steps: Foundational Research

Much foundational research is still needed before we can fully interpret developmental changes in pupil diameter or blink rate. For example, additional longitudinal studies tracking individuals between infancy, childhood, and adolescence are needed to characterize developmental changes in these measures, and how they relate to changes in cognitive performance. So far, studies have shown cross-sectional differences in tonic pupil dilation (Loewenfeld & Lowenstein, 1993; MacLachlan & Howland, 2002), the saccade system (e.g., Luna, Garver, Urban, Lazar, & Sweeney, 2004), and for blink rate (Zametkin et al., 1979; Cruz et al., 2011). But far less is known about systematic differences in task-evoked pupillary responses and blink rates. Only studies that examine these changes longitudinally will be able to characterize the precise developmental trajectory of the three ocular systems and to reveal individual differences. This research is necessary as a foundation for studies comparing age groups because differences in ocular responses need not necessarily arise from differences in cognitive processing, as which they are usually interpreted in cognitive studies. Only knowledge about the typical development of these responses will allow for an adequate interpretation of differences between age groups.

Longitudinal studies combining ocular measurements with cognitive tasks are needed to characterize the relationships between cognitive processing and changes in the underlying neurophysiology during development. This research is crucial for several reasons. First, the close links between the LC-NE system and pupil dilation and between the DA system and spontaneous blink rate have been established solely based on adult (and animal) work, and need to be replicated in children and adolescents. Second, several differences in ocular responses are conceivable between children and adults. These differences need to be characterized, explained, and related to cognitive processing. For example, the magnitude of the task-evoked pupillary response may change during development merely as a result of the dramatic changes in baseline pupil dilation; the same is true for spontaneous blink rate. It is important to study in what ways these changes, likely reflecting the maturation of the underlying neural system, are related to changes in cognitive processing. Third, the question needs to be answered as to how far developmental changes in the ocular measures are driven by age, pointing to a role of physiological maturation, versus how much they are driven by cognitive abilities, pointing to a role of experience. Very simple paradigms for which performance can be well-matched across ages (e.g., basic sensory discrimination paradigms, such as the auditory oddball task in Wetzel et al., 2016), should be used to answer this questions. Lastly, no non-invasive measures of the NE and DA neuromodulatory systems have been developed for use in humans yet (i.e., not requiring injection of radioactive isotopes, as is done for positron emission tomography (PET) studies of dopamine receptor binding). In the absence of such tools, pupil dilation and blink rate are our best available options for probing neurochemical underpinnings of cognition in pediatric populations. These methods can facilitate the study of the early development of abilities such as attention, working memory, cognitive control, decision making, and learning.

## 3.3 Future Directions: Application to the Study of Cognitive Development

The adult cognitive literature employing ocular measures has revealed a number of promising avenues for research, some of which we have reviewed above. Many of the used paradigms lend themselves to investigations in younger participants, shedding light on the development of the cognitive abilities in question. We will first focus on using pupillometry to study attention and using blink rate to study motivated behavior. After that, we will turn to more general topics and discuss how ocular measures can be used to identify cognitive processes that underlie developmental changes and how learning progress or the effects of interventions can be tracked using these measures.

One example of adult research that could reveal new aspects of neurocognitive development has focused on the influence of LC-NE activity on attention, as measured via pupillometry (Eldar, Cohen, & Niv, 2013). In this study, participants with different levels of LC-NE activity (or "neural gain", in the terms of the adaptive-gain theory; Aston-Jones and Cohen, 2005) were identified based on their task-evoked pupillary responses. Participants with the largest neural gain were found to show a strong correlation between attention during the task and their attentional predisposition. In other words, participants who were predisposed to attend to certain image features, independently of the task at hand, attended to these same features during the task. Participants with smaller neural gain, on the other hand, showed no such correlation, or even a negative one. These participants attended to any image features, independent of their own predisposition. This relationship reveals that high levels of neural gain led participants to rely heavily on their attentional predispositions, whereas lower levels led participants to relax these predispositions and to explore features more equally. The correlation between levels of neural gain and reliance on attentional predispositions was almost perfect ($r = 0.96$). Employing a similar task design in children could reveal if the LC-NE system is equally important for guiding attention during development. If the relationship were indeed found to be similarly strong, the pupillary response alone could be used as a quantitative measure of this aspect of attention, making obsolete verbal responses and thereby facilitating research on attention in even younger children, including preverbal infants.

The use of blink rate to study motivated behavior is another example of adult research for which translation into developmental studies seems promising. DA release in the striatum typically occurs in response to rewards and reward-predicting cues (Schultz, 2013). In reward-motivated behavior, the initial valuation of a reward or reward-predicting cue is thought to drive the allocation of effort to a task. The striatum is implicated in this stage of processing, which can occur without awareness of the signal or presence of the reward. In contrast, later reward processing stages can inform strategic decision-making and engage cortical areas (Bijleveld, Custers, & Aarts, 2012). Using blink rate as a proxy of striatal DA activity, Pas and colleagues (2014) found that individuals with higher tonic blink rate exerted more effort in a finger-tapping task when they were presented with cues that indicated higher pay-off of their work, but this effect only occurred in response to reward cues presented subliminally (i.e., an extremely fast and masked cue). Prior to this work, studying the neurobiological underpinnings of the initial phase of reward processing in humans was more difficult due to temporal and/or spatial constrains of neuroimaging methods (e.g., van Hell et al., 2010). Thus, blink rate can provide an invaluable way to examine reward processing. In particular, it could be used to characterize age-related differences at different stages of reward processing, which could inform current models and theories of decision-making during adolescence.

More broadly, ocular measures can be helpful in determining which of several candidate processes determine age-related differences in performance. If the experimental protocol achieves a separation in time of the processes in question, for example through sequential phases of encoding, retention, and recall in a memory experiment, the ocular responses for each can be assessed and compared independently. Such paradigms are useful in developmental research whenever it is necessary to pinpoint which cognitive processes are affected in a complex task, or to determine what the underlying reasons are for differences in behavior.

Another potential application of ocular measures is to track learning progressions and to specify the underlying mechanisms of behavioral change. For example, reduced cognitive effort associated with task performance (i.e., greater efficiency) could be evident in reduced task-evoked pupillary responses and blink rates. Reduced numbers of saccades would also suggest increased efficiency, and the same is true for fixation patterns that are more concentrated on the crucial elements of a task. In other words, quantitative changes in the ocular responses often reveal changes in cognitive efficiency, rather than qualitative changes in strategy. If an intervention led to changes in strategy, qualitatively different ocular response patterns would be expected, such as pupil dilations and blink rates whose intensities have shifted from some task components to others, or eye movements with qualitatively different patterns of saccades and fixations. Ocular measures are therefore a promising tool for determining whether a behavioral improvement reflects continuous or discontinuous, quantitative or qualitative change in underlying mechanisms.

### 3.4 Summary

Summing up, three different measures of the eyetracking toolbox, eye gaze, pupil dilation, and blink rate, have the promise of accessing crucial aspects of cognitive processing, such as attention, working memory, decision making, and cognitive control, across age groups. In this review, we have described the neural systems underlying pupil dilation and spontaneous blink rate, the LC-NE and DA systems, respectively. Amongst others, these systems play central roles in a number of complex cognitive functions, such as attention and working memory (LC-NE), and reward processing and cognitive control (DA). Building on a characterization of these neural systems, we next exemplified the range of cognitive studies in which these methods have been applied so far. Due to gaps in the literature in the use of these measures in the study of cognitive development, these sections focused mainly on adult work. The subsequent sections introduced a small number of selected examples from the developmental literature to show what kinds of analyses can be done with each measure and what results have been obtained in a developmental setting. We then explained the methodological details and considerations of each method.

In the final section, we compared eye gaze, pupil dilation, and blink rate in terms of appropriate task designs and underlying cognitive processes, and showed that these measures can be used as complementary measures of different aspects of cognitive processing. We also highlighted gaps in the current literature that need to be addressed to provide a solid foundation for developmental studies and studies comparing different age groups with these methods. Lastly, we pointed out several directions in which future research using ocular measures could advance the study of cognitive development. Research on attention and motivated behavior might be areas that will especially benefit from these methods, but the methods also have the potential to unveil aspects of neurophysiological processes underlying a much broader range of cognitive functions. Finally, we argued that qualitative and quantitative changes during learning and development could be differentiated with these measures. In closing, we expect that the eytracking toolbox – the measurement of eye gaze, pupil dilation, and blink rate – will yield novel insights about cognitive development over the coming years.

**Next Chapter:** Using the methodology discussed in this chapter, in Chapter 5 and 6, I provide empirical evidence about the mechanisms that support improvements in reasoning performance in young adults who underwent a real-world learning experience.

# Chapter 5

## Eye gaze patterns reveal how reasoning skills improve

Co-Author: Silvia A. Bunge[1, 2]

[1]Department of Psychology, University of California at Berkeley

[2]Helen Wills Neuroscience Institute, University of California at Berkeley

**Permission to include co-authored material:** Granted

**Acknowledgments:** We thank Kaplan, Inc., Dan Wittich, Adam Hampshire at CBS, Ariel Starr, Wei-Chun Wang, Monica Ellwood-Lowe, Martha Bawn, Ron Orellana, Carlos Castillo, Pooya Ganjali, Marissa Wong, Lara Taniguchi, and the study participants.

**Competing Interests:** Kaplan, Inc. provided the LSAT course materials and Cambridge Brain Science Inc., provided the online assessments used in the study. Aside from providing research materials, these companies had no role in the study design, data analysis, decision to publish, or preparation of the manuscript. B.G.C. and S.A.B. have no financial interest in these companies or any of their products.

**Funding:** National Science Foundation Graduate Research Fellowship (B.G.C.), the James S. McDonnell Foundation Scholar Award (S.A.B.), and the Jacobs Foundation Advanced Career Research Fellowship (S.A.B.).

## Abstract

Reasoning, our ability to solve novel problems, has been shown to improve as a result of learning experiences. However, the underlying mechanisms of change in this high-level cognitive ability are unclear. We hypothesized that possible mechanisms include improvements in the encoding, maintenance, and/or integration of relations among mental representations – i.e., relational thinking. Here, we used several eye gaze metrics that we developed to pinpoint learning mechanisms that underpin improved reasoning performance. We collected behavioral and eyetracking data from young adults who participated in a Law School Admission Test preparation course involving word-based reasoning problems or reading comprehension. The Reasoning group improved more than the Comprehension group on a composite measure of four visuospatial reasoning assessments. Both groups improved similarly on an eyetracking paradigm involving transitive inference problems, exhibiting faster response times while maintaining high accuracy levels; nevertheless, the Reasoning group exhibited a larger change than the Comprehension group on an ocular metric of relational thinking. Across the full sample, individual differences in response time reductions were associated with increased efficiency of relational thinking. Accounting for changes in visual search and a more specific measure of relational integration improved the prediction accuracy of the model, but changes in these two processes alone did not adequately explain behavioral improvements. These findings provide evidence of transfer of learning across different kinds of reasoning problems after completing a brief but intensive course. More broadly, the high temporal precision and rich derivable parameters of eyetracking make it a powerful approach for probing learning mechanisms.

**Keywords:** reasoning, plasticity, education, learning, eyetracking

# Introduction

Reasoning, the ability to solve novel problems, relies on multiple cognitive processes including various aspects of cognitive control and relational thinking (Goodwin & Johnson-Laird, 2006). Indeed, relational thinking is an essential component, as it allows us to form relational representations from mere percepts (Alexander, 2016). Solving reasoning problems, such those involving transitive inference, relies heavily on processes supported by relational thinking, including the ability to encode, maintain, and integrate mental relations (Goodwin & Johnson-Laird, 2006; Halford et al., 2010). Together, these processes allow us to identify patterns and solve novel problems, and are fundamental for human learning (Halford et al., 2010).

Prior research has demonstrated that reasoning can improve with targeted practice and increased task-specific expertise across the lifespan (Alexander, 2016; Bergman Nutley et al., 2011; Klauer & Phye, 2008; Knoll et al., 2016). However, it is still unknown which aspects of reasoning contribute to improved behavioral performance. Do people become more efficient at relational thinking with experience? To address this question, we leveraged the high temporal precision and rich derivable parameters of eyetracking to index cognitive processes that may support improvements in reasoning over time.

In earlier work, our lab demonstrated that young adults who underwent 100 hours of preparation for an exam that taxes relational reasoning (the Law School Admission Test, LSAT) showed improvements in reasoning performance and changes in the frontoparietal network (Mackey et al., 2013; Mackey, Miller Singley, Wendelken, & Bunge, 2015; Mackey et al., 2012). Compared to a passive control group, the LSAT group improved more in accuracy and speed on a test of transitive inference (Fig. 1A) that required the integration of novel visuospatial relations (Mackey et al., 2015; Wendelken & Bunge, 2009). Moreover, they showed a greater concomitant decrease in activation of dorsolateral prefrontal cortex (Mackey et al., 2015), a region broadly implicated in high-level cognition (e.g., Fuster, 2015). The LSAT group also showed changes in structural and resting-state functional connectivity of the frontoparietal network (Mackey et al., 2013, 2012), particularly between regions implicated in relational thinking (Krawczyk, 2012; Prado, Chadha, & Booth, 2011). Together, these findings provide evidence of experience-dependent brain plasticity as a result of practice with reasoning. However, these brain imaging results alone are insufficient to conclude which cognitive mechanisms were altered by the intervention (Constantinidis & Klingberg, 2016; Poldrack, 2015).

Candidate mechanisms that may underlie improvements in reasoning include the ability to identify relevant pieces of information (visual search) (Goodwin & Johnson-Laird, 2006), and relational thinking processes such as the encoding, maintenance, and integration of relations (Alexander, 2016; Halford, Wilson, & Phillips, 1998). Manipulating any of these elements have been shown to influence reasoning performance. For example, reasoning problems become harder as the number of relations one needs to integrate increases (Halford et al., 2010; Johnson-Laird, 2010). Additionally, people tend to make incorrect deductions when individual premises contain convoluted wording, as this hinders relational encoding (Goodwin & Johnson-Laird, 2006). Moreover, drawing attention to relevant relations or segmenting a complex task to facilitate focus on single relations can improve performance, even after controlling for other cognitive demands of the task (Duncan, Chylinski, Mitchell, & Bhandari, 2017). These examples illustrate the numerous factors that can influence reasoning ability.

In the present study, we sought to determine which cognitive processes, if any, are honed with experience – both as a result of reasoning instruction/practice and more generally from repeated experience with a test (i.e., a test-retest effect). To this end, we probed changes in patterns of eye movements on a reasoning task. In the ~7 seconds that it takes to solve one of the problems on this task, participants make ~23 eye fixations. As such, we posited that analyzing patterns of eye movements should be able to reveal more about the process of reasoning and mechanisms of learning than could accuracy, RTs, or fMRI activation.

Participants in this study performed a transitive inference task (Fig. 1A; adapted from Mackey et al., 2015; Wendelken & Bunge, 2009) while we collected eyetracking data, before and after they completed one of two online LSAT preparation courses developed by Kaplan, Inc. The Logic Games course focused on reasoning about novel problems, and the Reading Comprehension course on answering questions about passages of text (see Methods for sample problems for both sections of the LSAT). Our eyetracking task requires participants to jointly consider a subset of relevant visuospatial relations depicted by balance scales (see Figure 1). On the surface, this task bears no resemblance to the text-based problems in the LSAT curriculum. However, at a deeper level, both tax relational thinking.

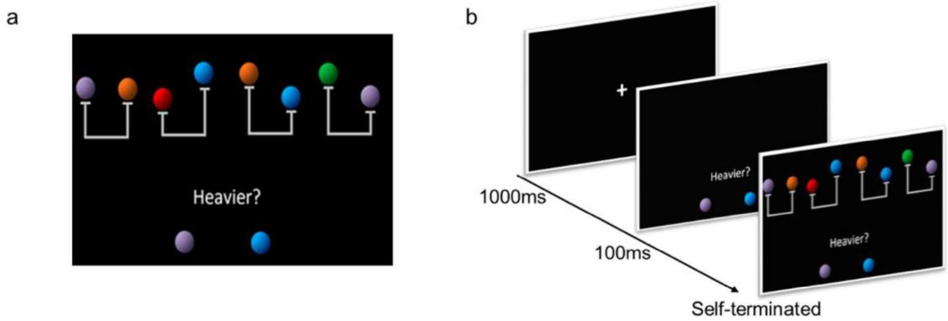


**Figure 1. Transitive inference task. (a)** Sample stimulus array, with four relations at the top and the question at the bottom. In this sample problem, participants had to encode that the blue ball was heavier than the orange one, and that the orange and purple balls were equally heavy, to determine that the blue ball was heavier than the purple one. **(b)** Eyetracking adaptation from (Mackey et al., 2015; Wendelken & Bunge, 2009): each trial began with a fixation cross in the center of the screen (1000ms) that cued participants to fixate on it, followed by the presentation of the question and target balls. After 100ms, four scales would appear, only two of which were relevant to the problem. A trial ended immediately after the participant pressed a button to indicate which of the two target balls was heavier.

We developed three gaze metrics to assess (1) visual search, (2) a broad measure of relational thinking encompassing encoding, maintenance, and integration of relations, and (3) a more specific measure of relational integration (see Table 1 and Methods for details). We based these metrics on findings from eyetracking studies that relate similar ocular behaviors to the processes of searching for task-relevant information (Rao, Zelinsky, Hayhoe, & Ballard, 2002), stimulus encoding and maintenance (Hannula et al., 2010), and relational integration (Thibaut & French, 2011; Vigneau et al., 2006).

We had initially sought to use eye gaze metrics to isolate three distinct stages of task performance: visual search, relational encoding, and relational integration (S3). However, examination of the eye gaze data (collapsed across groups and timepoints, Fig S2) did not support such clear-cut stages of processing (see also Goodwin & Johnson-Laird, 2006); rather, the data indicated a more gradual transition from visual search to relational integration. As such, we refined the planned gaze metrics to account for the reality of how participants solved the problems (see methods), and then tested for effects of group and timepoint.

**Table 1**. Gaze metrics indexing processes that may support improvements in reasoning

| | | Evidence supporting $H_1$<br>$BF_{10} \approx P(H_1 \mid \text{data}) / P(H_0 \mid \text{data})$ | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | $H_1 = POST < PRE$ | | | | | | $H_0 = G+T$<br>$H_1 = G \, x \, T$<br>$BF_{10}$<br>*(% error)* |
| | | Reasoning | | | Comprehension | | | |
| Cognitive Process | Gaze metric | PRE *Mdn [95%CI]* | POST *Mdn [95%CI]* | $BF_{10}$ *(%error)* | PRE *Mdn [95%CI]* | POST *Mdn [95%CI]* | $BF_{10}$ *(%error)* | |
| Visual Search | Decrease in the number of fixations on any scale before honing in on the relevant scales | 5.75 [4, 7] | 5 [4, 5] | 6.51* (±<0.00) | 5 [4, 5.75] | 4 [3, 5] | 0.67 (±<0.00) | 0.60 (±2.79) |
| Relational Thinking | Decrease in the total duration of fixations on relevant relations after honing in on the relevant scales | 2526.5 [1010.5, 3259.00] | 1122.5 [555.99, 1854.50] | 240.03*** (±<0.00) | 1635.75 [1069, 2256.25] | 1725 [1387.5, 1990.18] | 0.30$^{\bullet}$ (±0.02) | 3.66* (±1.98) |
| Relational Integration | Fewer saccades between the two relevant scales after honing in on the relevant scales | 3 [1.75, 3.75] | 2 [1, 3] | 3.15* (±<0.00) | 2 [2, 3] | 2.5 [2, 3] | 0.14$^{\bullet}$ (±<0.00) | 1.58 (±2.62) |

*Mdn[95%CI]: Median with 95% confidence intervals,* calculated with 1000 bootstrap iterations. $H_1 = POST < PRE$ assessed with Bayesian paired single-sided t-test. Interaction models (Group *x* Time) tested with Bayesian mixed regressions. Estimations made using BayesFactor's (Morey, Rouder, & Jamil, 2015) default Cauchy prior scale $r = \frac{1}{2}\sqrt{2}$ and prior uniform probability to the models. Refer to Table S1 for specification of the models and posterior odd estimates. Approximate classification scheme for the interpretation of Bayes factors from (Wagenmakers et al., 2017): *** Extreme evidence for $H_1$, *Moderate evidence for $H_1$,, $^{\bullet}$ Moderate evidence for $H_0$.

Our key predictions were that reasoning instruction/practice would be associated with improved reasoning performance and efficiency in relational thinking, including the more specific measure of integration. We tested these hypothesis with behavioral and gaze data from the transitive inference task. We also assessed behavioral improvements with a composite measure of four reasoning tests, to better characterize the generalizability of the intervention (Lindenberger et al., 2017; Lövdén, Bäckman, Lindenberger, Schaefer, & Schmiedek, 2010; Noack et al., 2014). The behavioral test battery additionally included assessments of working memory, planning, and selective attention, which we used to characterize the extent of transfer to untrained tasks. However, we did not anticipate far transfer to these measures, given limited evidence to date of far transfer of learning in adults (Au et al., 2015).

Finally, we undertook an exploratory analysis to understand the cognitive mechanisms that support test-retest improvements on the transitive inference task, and underlie individual differences in pre-test performance. Thus, we examined the relationship between relational thinking, integration, visual search, and behavior.

## Methods

### Ethics statement

The research was approved by the Committee for the Protection of Human Subjects at the University of California, Berkeley. Written informed consent was obtained from all participants.

### Participants and eligibility

We recruited college students planning to take the LSAT within one year. Inclusion criteria included being native English speakers, at least 18 years old, normal/corrected vision, and no history of psychiatric disorders, learning disabilities, or prior LSAT experience. Participants were assigned pseudo-randomly to study for one of these two sections of the LSAT, the Logic Games or the Reading Comprehension section. The first quarter of participants were assigned to a group at random, whereas we distributed the rest to match the groups on age, gender, reasoning, working memory, and LSAT performance (Table S1).

Ninety-five participants completed the pre-tests, and 49 completed the LSAT course and post-tests. We excluded two of these participants because they failed to study for their assigned course. Participants in our final sample did not differ from those who only completed one timepoint on either cognitive performance or demographic variables. The final sample who prepared for the Logic Games section of the LSAT included 23 students (14 Females, mean age 21.55). The final sample who prepared for the Reading Comprehension section of the LSAT included 24 students (13 females, mean age 21.88). Levels of attrition did not differ significantly between the groups ($X^2 = 0.01$, $p = 0.93$). For analyses involving the transitive inference task, we excluded two subjects from each group for having more than 60% of trials missing valid fixation data, and one subject from the Reasoning group for having performance below chance levels (20% accuracy, chance was 50%).

### Summary of procedures

Before and after studying for the LSAT courses, participants completed a battery of nine online cognitive assessments (Hampshire, Highfield, Parkin, & Owen, 2012), followed by an in-person testing session. Participants were blind to their LSAT group at pre-test, and the experimenters carrying out the testing sessions were blind to the group assignment at both timepoints.

During the lab sessions, we recorded gaze data from participants while they completed a transitive inference task, followed by two tests of inductive reasoning. Data from the transitive inference task is the subject of the current investigation. After finishing the eyetracking tasks, participants completed a standardized test of reasoning termed Analysis Synthesis (Woodcock-

Johnson Battery III; Woodcock, McGrew, & Mather, 2001), LSAT sample problems, and a survey. The survey included demographic and ocular health questions, questions regarding prior experience with the LSAT, and at post-test, questions about the participant's experience with their LSAT course. The order of tests was the same at both timepoints.

## LSAT courses

Participants studied for either the Logic Games or Reading Comprehension section of the LSAT with a commercially available online course (Kaplan, Inc.) for 6 weeks. During the intervention, students only had access to the instructional material related to their assigned section, and we monitored their progress in the course. We requested participants to 1) study only for the LSAT section we assigned to them, 2) complete all six lessons of the course within 7 weeks (~one lesson/week), and 3) space their practice (i.e., study every other day, three times per week). The latter was important because prior work has established that spacing practice promotes learning (Greene, 1989) and transfer effects (Wang et al., 2014). We chose these practice intervals so that students could incorporate their LSAT courses more easily with their typical school schedules. Participants reported having complied with these instructions, and that they completed on average, one lesson per week (range = 0.5-2 lessons) and studied their course for 24 hours (range = 4-24h). Both groups reported similar studying times.

The Logic Games section involves solving word problems that contain many rules that must be integrated to find the correct answer (sample problems: https://www.lsac.org/jd/lsat/prep/analytical-reasoning). The preparatory course for this section instructed on strategies such as organizing relational information into sketches to minimize the amount of information one needs to remember, as well as to facilitate deductions, rule abstractions, and correct rule application.

The Reading Comprehension section involves reading long passages and answering multiple choice questions based on relevant information in the passages (sample problems: https://www.lsac.org/jd/lsat/prep/reading-comprehension). The preparatory course for this section involved learning strategic reading techniques, such as finding keywords based on the passage questions and annotating main ideas on the passages to minimize working memory demands.

The courses were similar in critical ways. Both courses included six lessons, each consisting of online videos and homework practice problems designed to help improve timing and increase mastery with different question types. Both courses featured the same instructors in the online videos, who explained problem-solving strategies and had students practice those skills with real LSAT problems. Participants in the two groups found their respective courses similarly effective and enjoyable (Table S2).

We measured the effectiveness of the LSAT courses with short Logic Games,and Reading Comprehension problem sets that participants completed in the lab. We found little evidence of the effectiveness of the mini-courses in improving performance on either section (S1). However, our participants had pre-test scores on a full practice LSAT exam (Table S1) comparable to the scores on the first practice test attained by the LSAT students in our lab's previous study (Mackey et al., 2015).

### Eyetracking apparatus and procedures

We recorded binocular gaze data from participants completing a transitive inference task using Tobii T120 Eye Tracker (17-inch monitor, 1280 ×1024 pixel resolution). We sampled at a temporal resolution of 120Hz, with participants sitting at 60 cm from the eyetracker camera. We took several precautions to collect high-quality ocular data following recommendations from (Holmqvist et al., 2011). Furthermore, participants reported that they did not suffer from medical conditions or used medication that could affect ocular behaviors. We used Presentation® software (v. 18.0, Neurobehavioral Systems, Inc.) to present the task stimuli and the Tobii Eye Tracker Extension for Presentation v1.1 (Martin et al., 2007) to synchronize the timing of the stimulus presentation and ocular events.

### Transitive Inference Task

In the transitive inference task (adapted from Mackey et al., 2015; Wendelken & Bunge, 2009; Fig.1), participants see four balance scales, each one with two color balls. Based on the relations shown by the scales, participants needed to infer the relative weights of two target balls. To solve the problems correctly, it was necessary to integrate the relationship shown by two of the four scales (i.e., the relevant scales). Participants completed 60 of these problems, divided into two blocks of 30 trials. We recalibrated the eyetracker during the short break between blocks.

We minimized potential confounds in gaze patterns by controlling for features that could impact visual saliency and subjects' expectations as to where the relevant scales were likely to appear and which balls were likely to be relevant. We changed the position of the relevant scales across trials, and the program selected the color of the five balls at random from a set of six colors, which were all matched in luminance. Additionally, we biased the participant's first fixation to the question area by first presenting the question alone, and then adding the four scales (see trial sequence in Fig 1.B).

**Behavioral outcome measures.** We examined changes in response times (RTs) and accuracy (proportion of trials answered correctly). Performance did not vary as a function of the spatial arrangement of the scales (e.g., the position of relevant scales) or the number of scales showing inequalities (Fig S1). Thus, we did not include these factors in our analyses in favor of maximizing the statistical power to assses our hypotheses.

Given that pre-test RTs were highly positively skewed ($sk = 4.55$), we trimmed outlier trials falling on the long end of the tail (i.e., Q3+1.5*IQR) to minimize bias in our gaze analysis that could result from including the highly variable fixation durations that could occur on these atypically long trials. Outlier trials were identified separately by subject, timepoint, and block, to retain individual differences in performance. Approximately 5% of trials were trimmed due to outlier RTs from each group per timepoint.

**Gaze preprocessing and outcome measures.** We used custom scripts written in Python v3.6 to preprocess and calculate gaze outcome metrics.

We classified gaze data into fixations using a standard dispersion-based algorithm adapted from (Salvucci & Goldberg, 2000), allowing a maximum dispersion of 35px over a 100ms window (see details in S2). Participants had a median of 22 fixations on correct trials. Our analysis included only trials with at least three valid fixations, under the assumption that this is the minimum number of fixations needed to solve the problem, with a maximum of 64 fixations (i.e., Q3 + 1.5*IQR) to minimize the bias that those outlier trials could induce.

We assigned an area of interest (AOI) label to the fixations. The AOIs included each of the four scales (two relevant and two irrelevant scales) and the area where the target balls and question appeared. We used these labeled fixations to calculate the number of gaze transitions between different AOIs. For instance, a fixation on "Relevant Scale 1" followed by a fixation on "Relevant Scale 2" was coded as one transition between the relevant scales. We refer to these events as transitions because we were primarily concerned with measuring how often fixations shifted between two different scales; we ignored, at most, one fixation that may have occurred elsewhere between those two target fixations.

We used the transitions and fixation data from each trial to derive three gaze outcome measures (Table 1), informed by an analysis of fixation sequences performed across groups and timepoints (Fig S2). To compute the gaze metrics, we first marked the point when it became more probable that a participant had honed in on the relevant scales during a trial. For each trial, and on an individual subject basis, we measured that point in the trial by calculating the empirical probability that the number of fixations on irrelevant scales was below chance (25%), and that the number of fixations on relevant scales was greater than chance. We estimated these probabilities with a sliding window that evaluated 20% of the fixations at once (min size 4, max size 8 fixations). This approach enabled us to capture a common pattern of fixations (Fig.2), whereby participants began to preferentially fixate on the relevant scales after a certain point in the trial. Accordingly, the visual search metric constitutes the number of fixations the participant made on any scale prior to that point, and we indexed relational thinking as the duration of fixations on relevant scales occurring after that point. We additionally computed a more conservative metric of relational integration as the number of saccades between the two relevant scales.

### Composite reasoning measure and other transfer tasks

Three subtests included in the composite reasoning measure (Table 2) were part of a larger battery of nine online assessments, which included tests of selective attention, planning, and working memory (Table S4). These tests were developed by the Cambridge Brain Sciences Laboratory (http://www.cambridgebrainsciences.com), as an online adaptation of assessments designed and validated at the Medical Research Council Cognition and Brain Sciences Unit (Hampshire et al., 2012; Owen et al., 2010).

Task difficulty in all the assessments was adaptive as a function of performance. Performance metrics differed between the tasks (e.g., a maximum level achieved vs. total correct responses), so we standardized the scores after removing outlier scores (i.e., scores that deviated more than 3 *S.D.* away from the grand pre-test mean). Using this normalized dataset, we created composite measures of reasoning, planning, and working memory by averaging performance across related assessments. Composite measures provide a robust test of trasnfer (Lindenberger et al., 2017; Lövdén et al., 2010) and help minimize the number of statistical tests necessary. We derived these composite measures with a theory-driven approach, given that factor analytic methods were not appropriate for our sample size. For the reasoning measure, we averaged the standardized scores from the Analogical Reasoning, Object Reasoning, and Odd One Out tests, as well as the Analysis Synthesis test administered in the lab.

### Statistical Analysis

We used Bayesian models to quantify the strength of evidence supporting the model that tested a given hypothesis in question, as described in the results section. For all analyses, we used participant's median scores on the measure of interest, and uniform distribution of prior probabilities with default Cauchy prior scales from the BayesFactor R package (Morey et al., 2015). We performed all analyses in Python 3.6 and R 3.2.

**Supplementary materials:** all supplementary information referenced in this chapter can be found in Appendix B.

## Results

### Improvements related to targeted reasoning instruction/practice

We quantified evidence in support of our hypotheses with Bayesian tests, permitting us to assess how likely our data is to support one model versus another using the Bayes Factor ($BF_{10}$), and thus also quantify the strength for the null hypothesis (Wagenmakers et al., 2017). We used Bayesian single sided t-tests to gauge support for the prediction that the Reasoning group would improve in the behavioral and gaze metrics. We followed these tests with Bayesian mixed regressions to assess the probability that these changes could be attributed to reasoning practice beyond test-retest alone or subject variance. As such, we quantified the strength of evidence in favor of including the Group *x* Time term relative to a model containing both main effects. We modeled subject variance as a random nuisance factor, but the model design is otherwise equivalent to a 2x2 repeated-measures ANOVA. We report $BF_{10}$ (see Table S3 for detailed output), and interpret this metric in accordance with prior work (Wagenmakers et al., 2017): $BF_{10} > 1$: data provide positive evidence for the hypothesis, $BF_{10} > 3$: moderate evidence, $BF_{10} > 10$: strong evidence. The inverse applies for the null hypothesis ($1/BF_{10}$).

**Transfer to the composite reasoning metric.** The Reasoning group showed approximately a 22% improvement on this metric; by contrast, there was no evidence that performance of the Comprehension group changed between timepoints (Figure 2; Table 2). The Group *x* Time interaction model also received strong support. Thus, according to the Bayesian analysis, there is strong evidence that the Reasoning group improved on the composite of four measures of reasoning; this was not the case for the Comprehension group.

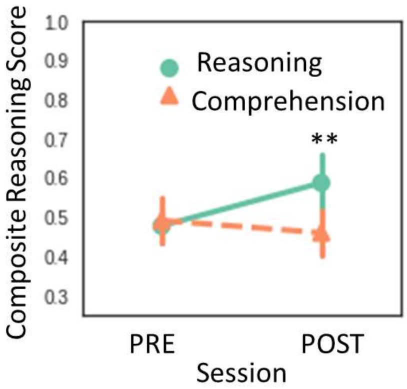


**Figure 2. Performance on reasoning assessments.** Scaled score on the composite measure of reasoning (y-axis), before and after (x-axis) each group completed their LSAT course. Error bars are 95% C.I estimated with 5000 bootstrap iterations. **Strong evidence that the Reasoning group showed greater improvements across timepoints. See Table 2 for detailed statistics.

**Table 2**. Transfer to Composite Reasoning Measure

| Subtest included | Description of subtest | Evidence supporting $H_1$<br>$BF_{10} \approx P(H_1 \mid \text{data}) / P(H_0 \mid \text{data})$ | | |
|---|---|---|---|---|
| | | $H_1 = POST < PRE$<br>$BF_{10}$ (~ % error) | | $H_0 = G + T$<br>$H_1 = G \times T$<br>$BF_{10}$<br>(% error) |
| | | Reasoning | Comprehension | |
| Odd One Out | Infer rules that relate object features to identify a deviant object among 9 choices | 20.78 **<br>(±<0.00) | 0.63<br>(±<0.00) | 15.06**<br>(±3.04) |
| Object Reasoning | Decide whether four 2x2 matrices containing geometrical patterns form a sequence | | | |
| Analogical Reasoning | Apply the rule governing the relationship between three objects to a new set of objects | | | |
| Analysis Synthesis | Solve logical puzzles involving color codes representing symbolic rules | | | |

$H_1 = POST < PRE$ assessed with Bayesian single-sided paired t-test. Interaction models (Group *x* Time) tested with Bayesian mixed regressions. Estimations made using BayesFactor's (Morey et al., 2015) default Cauchy prior scale $r = \frac{1}{2}\sqrt{2}$ and prior uniform probabilities. See Table S1 for model specification & posterior odd estimates. Approximate classification scheme for the interpretation of Bayes factors from (Wagenmakers et al., 2017): ** Strong evidence for $H_1$

We also tested for transfer to the measures of working memory, planning, selective attention, and verbal fluency. There was no evidence that performance on these measures changed between timepoints in either group (Table S4). Based on these results, there was evidence of transfer from the LSAT Logic Games course to a composite score of four measures of reasoning, but little evidence of far transfer from the study materials to other cognitive domains.

**Transfer to the Transitive Inference Task.** Given that accuracy was at ceiling already at pre-test (Fig. *S1*) for this sample, unlike the sample in our prior study (Mackey et al., 2015), we focused exclusively on response times (RTs) on correct problems as a measure of performance. The data provided strong evidence in support of the hypothesis that the Reasoning group would become faster at accurately solving the problems between timepoints ($BF_{10} \approx 27.41 \pm < 0.00$, $\Delta \approx 18\%$). There was positive, albeit weaker, evidence that the Comprehension group also improved ($BF_{10} \approx 2.18 \pm < 0.00$). Although the data provided support for the model containing the Group *x* Time term ($BF_{10} \approx 8.30 \pm 4.48$), the strongest model included only the effect of Time ($BF_{10} \approx 39.24 \pm 2.41$). Thus, although the evidence of improvement was stronger for the Reasoning group, both groups got faster at solving the task between timepoints (Fig. 3A).

**Gaze metrics.** We predicted that reasoning instruction would lead to improved efficiency of relational thinking, including changes in relational integration. For the Reasoning group, the gaze data provided moderate support for the model stipulating changes in the selective measure of relational integration, and extreme evidence in favor of the model testing improvements in the broader measure of relational thinking (for these and subsequent results, see Table 1; Fig. 3.B-D). By contrast, for the Comprehension group, there was moderate evidence in favor of the null model of both metrics, suggesting that their pre- and post-test scores are comparable. When considering whether the changes in relational thinking in the Reasoning group were greater than in the Comprehension group, we find that the data are 3.66 times more likely under the model including the Group *x* Time term compared to the model testing only the main effects. Indeed, the interaction model was best supported by the relational thinking data ($BF_{10} \approx 7.27 \pm 1.51$). Conversely, there was moderate evidence against including the interaction term to model the relational integration data ($BF_{10} \approx 0.17 \pm 2.32$). Thus, there is moderate evidence that reasoning instruction/practice led to improved efficiency of relational thinking.

We also tested for improved efficiency of visual search. For the Reasoning group, the data provided moderate support for the model stipulating increased search efficiency; this was not the case for the Comprehension group. However, the data did not support the Group *x* Time model: rather, there was moderate support for the main effect of Time ($BF_{10} \approx 4.01 \pm 1.14$). These results suggest that changes in visual search were likely due to individual differences, test-retest effects, or a general effect of being exposed to an intervention, rather than specifically an effect of reasoning instruction/practice.

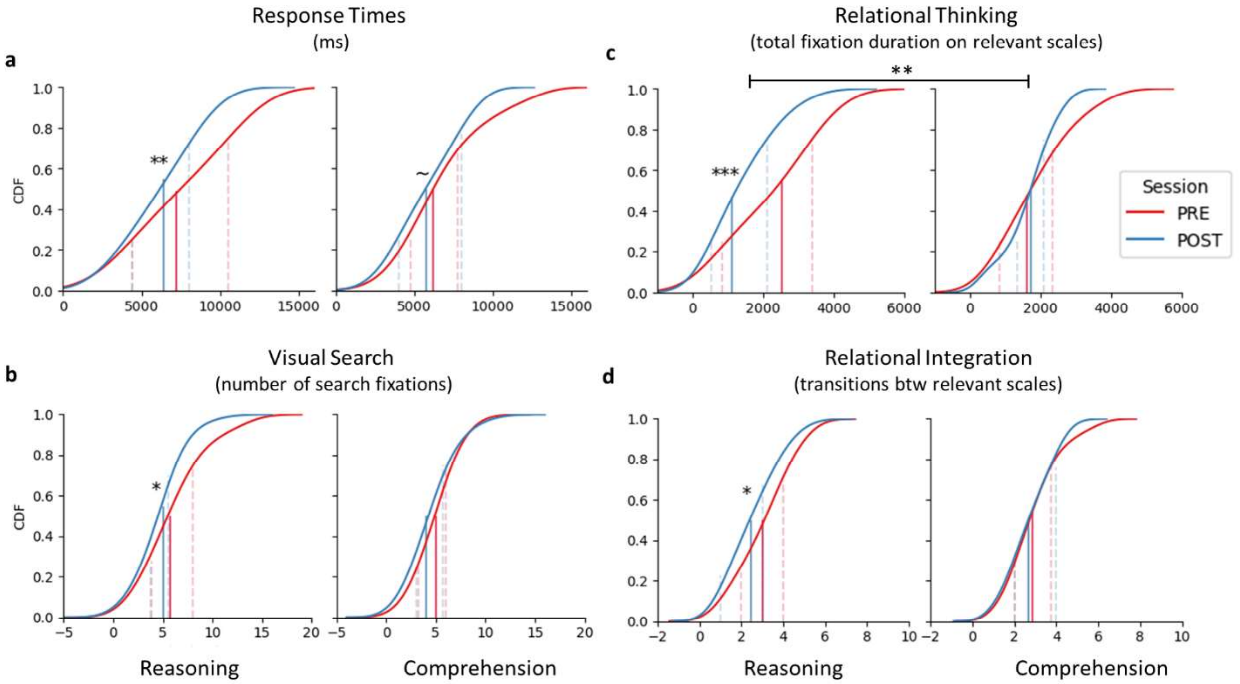


**Figure 3. Performance and gaze metrics from the transitive inference task.** Cumulative distribution functions (CDF) on each measure, for each group (left panels = Reasoning group) and timepoint (blue functions = post-test scores). Vertical lines denote: medians (solid), and 25th or 75th percentiles (dotted lines left or right to the median, respectively). Classification scheme to interpret Bayes factors (*BF*$_{10}$) used to quantify the strength of evidence in support of the models testing: improvements across timepoints (i.e., POST < PRE), and differential improvements between the groups (i.e., Group *x* Time): ***Extreme, **Strong, *Moderate, ~Anecdotal. Refer to Table 1 for detailed statistics.

## Individual differences in performance and test-retest improvements on the transitive inference task across the full sample

We first assessed the relationships among the gaze metrics at pre-test with Kendall's tau Bayesian correlations. At pre-test, relational thinking was strongly correlated with both visual search ($\tau$ =0.40, $BF_{10}$ ≈161.67 ) and relational integration ($\tau$ =0.64, $BF_{10}$ ≈4.742e +6), but the latter two metrics were only moderately correlated with each other ($\tau$ =0.29, $BF_{10}$ ≈6.41). Similarly, change in relational thinking correlated strongly with changes in relational integration ($\tau$ =0.53, $BF_{10}$ ≈24534.15), and moderately with changes in visual search ($\tau$ =0.29, $BF_{10}$ ≈6.61); by contrast, there was no evidence that changes in visual search and relational integration were correlated with one another ($\tau$ =0.19, $BF_{10}$ ≈0.91). These results provide evidence that greater efficiency of relational thinking was related to both visual search and relational integration, which in turn were separable components of relational reasoning.

Next, we used Bayesian regression models to determine which gaze metric(s) could best explain changes in RTs between timepoints across the full sample. The models included LSAT group as a nuisance variable and RT changes as a dependent variable. The predictors included one or more of the gaze metrics. We found the strongest evidence in favor of a model that included changes in relational integration, relational thinking, and visual search as predictors of change in RTs ($BF_{10} \approx 28.05 \pm 0.79$). Together, these metrics accounted for ~35% of the variance in RT reduction. In simpler models testing individual gaze predictors of changes in RTs, there was strong evidence in favor of relational thinking ($BF_{10} \approx 20.42 \pm 0.79$), which accounted for ~23% of the variance in RT reductions. By contrast, there was mild evidence for visual search ($BF_{10} \approx 2.81 \pm 0.80\%$, $R^2 = 0.14$), and no evidence in favor of the more specific measure of relational integration ($BF_{10} \approx 0.55 \pm 0.89$). Overall, these results point to relational thinking as a key facet of practice-related improvements in transitive inference task performance.

## Discussion

Which cognitive mechanisms underlie improvements in relational reasoning? We sought to address this question by examining improvements related to targeted reasoning instruction and practice with the Logic Games section of the LSAT, and general test-retest improvements across two timepoints by also considering the changes in the group who prepared for the Reading Comprehension section of the LSAT. Our key prediction was that practicing solving word-based logic problems would lead to improved performance on visuospatial tests of reasoning, as well as improved efficiency of relational thinking on a visuospatial transitive inference task, as measured via gaze metrics. We additionally tested whether either group exhibited increased efficiency of visual search.

We found evidence that reasoning instruction/practice led to improved performance on a composite of four measures of reasoning. On the surface, these measures of reasoning did not resemble the LSAT problems, but rather shared a deeper commonality of demands on relational thinking. Thus, these results provide evidence of moderate transfer from one type of reasoning practice to other reasoning tests.

Additionally, we found evidence that reasoning instruction/practice led to increased efficiency on our ocular measure of relational thinking. By contrast with the relational thinking metric, there was no compelling evidence that the changes in visual search or the conservative metric of relational integration could be attributed specifically to the reasoning intervention. Thus, we conclude that reasoning instruction/practice predominantly honed the ability to encode and maintain several mental relations in mind.

Although reasoning/instruction affected relational thinking on the transitive inference task, it did not yield a benefit in terms of behavioral changes on this task over and above a test-retest effect and the gains observed in the active control group. This discrepancy illustrates the idea that gaze metrics can pinpoint changes in specific cognitive processes even if the behavioral measures administered in the study are insufficiently sensitive. Similar arguments have been made with regard to brain imaging studies examining the effects of an intervention (Neville et al., 2013) or predicting future behavior (Gabrieli, Ghosh, & Whitfield-Gabrieli, 2015).

Having found evidence that the Reasoning group improved on reasoning measures, we sought to characterize the extent of transfer to other cognitive domains. In our prior study (Mackey et al., 2015), we did not observe transfer of reasoning instruction/practice to individual measures of matrix reasoning, rule induction, working memory, or processing speed. Here, we examined transfer to composite measures of various cognitive abilities, rather than individual tests, as a more robust test of transfer (Lövdén et al., 2010). We found moderate evidence in favor of the null hypothesis – i.e., no change – for measures of planning, working memory, selective attention, and verbal comprehension. Taking together the results from both studies, there is – as predicted based on prior intervention studies (Au et al., 2015; Melby-Lervåg & Hulme, 2013) – no evidence of far transfer from LSAT practice.

Finally, we adopted an individual differences approach to understand the processes that support test-retest improvements and performance on the transitive inference task. At pre-test, relational thinking was strongly correlated with both visual search and relational integration, but these two metrics were only weakly correlated with each other. This pattern of results was also obtained for correlations among the change scores for these metrics. These findings suggest that there could be temporal overlap between visual search and the early stages of relational thinking (i.e., initial encoding of scales), whereas relational integration may overlap with the later stages (i.e., maintenance).

Across the entire sample, we found evidence that RT reductions were associated with improved efficiency of relational thinking and visual search, but that relational thinking was likely the strongest driver of change. This finding, along with the pattern of correlations among gaze metrics, suggests that these gaze metrics capture at least partially separable cognitive processes – and that each contributes differentially to improved performance. Indeed, even when accounting for improved attentional control underlying visual search, changes in relational thinking are still a critical predictor of improved reasoning performance.

Although this study leverages eyetracking measures in a novel way to provide insights regarding learning mechanisms, there are several limitations to consider. First, while we have evidence that practicing Logic Games is associated with gains in other measures of reasoning, we lack strong evidence that Logic Games performance itself improved after the 6-week online course, as measured via one brief (8-minute) problem set administered at pre-test and two at post-test. This outcome contrasts with the improvements we found in our lab's prior study (Mackey et al., 2015), which differed from the current study in multiple ways. In particular, changes in LSAT performance in the previous study were assessed with full-length practice exams, which included 4 problem sets (35 minutes) for Logic Games as well as for Reading Comprehension. Although the problems we had selected are considered of medium difficulty and test common question types, the strategies taught in the LSAT course may not have been particularly useful for solving the specific problems we selected, or perhaps students needed additional instruction on how to apply those strategies effectively. Alternatively, the online course format – while ideal from an experimental standpoint, as it enabled us to compare two separate but similarly structured courses – may not have been an ideal learning platform. However, participants rated both courses as effective. Finally there may have been a synergistic effect in the previous study of studying for all sections of the LSAT together (Logic Games, Analytical Reasoning, and Reading Comprehension), and spreading the course over 3 months as opposed to 6 weeks. However, despite the lack of improvement on our brief Logic Games assessment, we contend that we can meaningfully assess effects of this experience on other assessments that tap overlapping skills.

Second, pre-test accuracy on the transitive inference task was at ceiling, in contrast with our prior study (Mackey et al., 2015). This difference likely reflects differing sample characteristics. The task was sensitive to RTs in this study, but the two groups sped up to a similar degree. We can only speculate that if the task had been more difficult for these participants, we would have had an opportunity to observe a differential effect of Reasoning and Comprehension courses on accuracy.

A final limitation is that recruitment and retention were challenging. The study required students who were inexperienced with the LSAT and were willing to commit to studying for only one section – a requirement that likely dissuaded students who sought to take the LSAT immediately. Additionally, the study required a serious time commitment for undergraduates who already had a full course load. However, considering that there were similar levels of attrition in the Reasoning and the Comprehension groups, we are still able to draw meaningful conclusions about the effects of reasoning practice.

To conclude, our study highlights the utility of eyetracking for probing the mechanisms underlying real-world learning. The gaze measures used in our study revealed that changes in relational thinking contributed to improved reasoning performance, beyond changes in supporting attentional processes. Additionally, the high temporal resolution of the eyetracker provides a more detailed window into the series of rapid computations and highly interactive processes that underlie reasoning (Duncan et al., 2017) than is possible with neuroimaging or behavioral methods alone. Beyond elucidating mechanisms of plasticity, then, this study also provides a window into the thought processes that unfold during reasoning. In future research, the combined use of eyetracking with behavioral and neuroimaging methods could provide unique insights into the brain mechanisms that support cognitive functioning and learning, and sources of individual differences therein.

**Next chapter:** given that relational thinking was the mechanisms impacted by reasoning instruction, we next examine the degree of transfer to another domain in reasoning that also relies in the same cognitive process but in the service of rule induction. In the next chapter, we examine the degree of transfer from practicing reasoning in a context that primarily taxes deductive reasoning to performance in a rule induction task, which we would consider a further degree of transfer than what was documented in Chapter 5. Furthermore, we use metrics of pupil dilation to characterize whether and how the learning experience influences how rule induction unfolds during problem-solving.

# Chapter 6

# Phasic pupillary responses reveal insights into practice-related changes in inductive reasoning

Co-Authors: Maria K. Eckstein[1], Ariel Starr[1], & Silvia A. Bunge[1, 2]

[1]Department of Psychology, University of California at Berkeley

[2]Helen Wills Neuroscience Institute, University of California at Berkeley

**Permission to include co-authored material:** Granted

**Acknowledgments:** This research was supported by the National Science Foundation Graduate Research Fellowship (B.G.C.), the German academic exchange service (Deutscher Akademischer Austauschdienst) (M.K.E), the National Research Service Award postdoctoral fellowship (A.S), the James S. McDonnell Foundation Scholar Award (S.A.B.), and the Jacobs Foundation Advanced Career Research Fellowship (S.A.B.). We thank Kaplan, Inc., Dan Wittich, Adam Hampshire, Martha Bawn, Ron Orellana, Carlos Castillo, Pooya Ganjali, Marissa Wong, Lara Taniguchi, and the study participants.

**Disclosure Statement:** Kaplan, Inc. provided the LSAT course materials used in the study. Aside from providing research materials, these companies had no role in the study design, data analysis, decision to publish, or preparation of the manuscript. None of the authors have a financial interest in the company or in any of its products.

## Abstract

Relational reasoning has been shown to improve following learning experiences that emphasize either deductive or inductive reasoning. Even though both inductive and deductive reasoning rely on shared cognitive and neural mechanisms that support relational reasoning, the evidence of transfer of learning between these domains is scant. Here, we examine transfer from practicing reasoning in a domain that mainly emphasized deductive reasoning over a short time-frame to performance on a rule induction task named "SET" (Eckstein, Starr, & Bunge, under review). We hypothesized that completing a 6-week preparatory course for the Logic Games section of the Law School Admission Test (LSAT) would lead to greater improvements in the rule induction task, and influence how rule inference processes unfold during problem-solving. Despite prior evidence that the Reasoning group improved in relational thinking efficiency during a transitive inference task and had improved performance on a composite measure of four tests of relational reasoning (Guerra-Carrillo & Bunge, under review), we found moderate evidence against the hypothesized transfer effect to this rule induction task. However, accuracy was already at ceiling prior to training. By contrast, the pupillary measures provided evidence that this group engaged in active rule inference to a greater degree at post-test, whereas there was no evidence of this change in the active control group. Although there was no evidence of this change in the active control group, the changes in the Reasoning group were likely due to test-retest effects or individual differences. Thus, we consider these results as evidence of practice-related improvements in the process of active rule induction. We postulate that the pupillary metrics could reflect ongoing changes over the course of an intervention that either accompany or precede changes in behavior.

**Keywords**: pupil dilation; rule inference; inductive reasoning; transfer; learning

## Introduction

Relational reasoning, our ability to derive meaningful relations between mental representations, is thought to be malleable – even in adulthood (for review, see Alexander, 2016). Considering the importance of reasoning in human cognition (Halford, Wilson, & Phillips, 2010), several studies have aimed to promote transfer to reasoning performance from a variety of learning experiences (Dumas & Schmidt, 2015; Gamino et al., 2014; Owen et al., 2010; Schmiedek, Lövdén, & Lindenberger, 2010). The most consistent evidence of positive transfer comes from studies involving targeted reasoning instruction and practice in both healthy (Bergman Nutley et al., 2011; Klauer, Meiser, & Naumer, 2000; Klauer & Phye, 2008; Mackey, Hill, Stone, & Bunge, 2011) and clinical populations (Chapman & Mudar, 2014; Christoforides, Spanoudis, & Demetriou, 2016; Han, Davis, Chapman, & Krawczyk, 2017). A key motivation in this line of work has been for the gains experienced during an intervention to support relational reasoning in a variety of novel contexts and tasks, but this type of broad transfer has not always occurred (Daniel & Klaczynski, 2006; Roth-Van Der Werf, Resing, & Slenders, 2002). Thus, an important question is, what are the boundaries of transfer in relational reasoning?

In the general context of cognitive plasticity, it has been argued that the degree of transfer depends heavily on the extent to which the trained and transfer tasks rely on shared neural and cognitive mechanisms – in particular, that performance on the transfer task depends on the cognitive processes taxed by the trained tasks (Lindenberger et al., 2017; Lustig, Shah, Seidler, & Reuter-Lorenz, 2009). There has been empirical evidence in support of this argument from interventions targeting higher-cognitive skills, such as working memory (Constantinidis & Klingberg, 2016). Work from our lab has also provided support for this argument in the context of relational reasoning with findings from two separate sets of studies, whereby we probed changes in the cognitive and neural mechanisms that support relational reasoning following targeted reasoning instruction and practice (Guerra-Carrillo & Bunge, under review; Mackey, Miller Singley, & Bunge, 2013; Mackey, Miller Singley, Wendelken, & Bunge, 2015; Mackey, Whitaker, & Bunge, 2012).

In one study, young adults who prepared for the Logic Games section of the Law School Admission Test (LSAT) with a 6-week online course involving approximately 33 hours of instruction and practice improved in relational thinking efficiency, as indexed by gaze metrics on an eyetracking transitive inference task (Guerra-Carrillo & Bunge, under review). The transitive inference task engages various aspects of relational thinking, including the ability to encode, maintain, and integrate visuospatial relations in the service of deducing the ordering of the stimuli (Wendelken & Bunge, 2009). We have postulated that these same deductive processes are emphasized in the Logic Games course, wherein students practice ordering and integrating verbal relations according to a set of given rules (Mackey et al., 2015). Compared to the active control group, who prepared for the Reading Comprehension section of the exam by completing a well-matched online course, the Reasoning group also showed greater gains on a composite measure of four tests of relational reasoning, which included tests of deduction and induction[2].

---

[2] Composite score included Analysis Synthesis from the Woodcock Johnson III (Woodcock, McGrew, & Mather, 2001), and three tests from Cambridge Brain Science online testing platform (Hampshire, Highfield, Parkin, & Owen, 2012): Odd one Out, Object Reasoning, and Analogical Reasoning (for details see Guerra-Carrillo & Bunge, under review)

We observed no evidence of transfer in either group to tests of working memory, planning, or selective attention. These findings suggest that transfer was limited to tests of reasoning, and provide evidence of the malleability of processes that underlie relational reasoning.

In a previous set of studies involving students who completed a 3-month course including all sections of the LSAT, in comparison with a well-matched passive control group, the LSAT group showed improved performance on the fMRI version of the transitive inference task (Mackey et al., 2015), as well as change in structural connectivity (Mackey et al., 2012a) and functional connectivity (Mackey et al., 2013) within a lateral fronto-parietal network that has been implicated in reasoning (Krawczyk, 2012; Prado et al., 2011). In that study, there was no transfer from LSAT practice to test of working memory and processing speed, and null results in regards to two tests of inductive reasoning. However, there were methodological issues with both of the inductive reasoning tasks, which rendered the null results ambiguous (see discussion in Mackey et al., 2015). Taken together, the results from all these prior studies provide evidence of moderate transfer from one type of reasoning practice to other tests of reasoning, because the measures did not resemble the LSAT problems at a surface-level (visuospatial puzzles as compared with word problems), but they all emphasized relational reasoning in some manner.

Considering the hypothesis that transfer can occur between domains with shared cognitive mechanisms, we could expect broader transfer between different types of context that rely on relational reasoning, such as between inductive and deductive reasoning. Indeed, both domains rely on shared cognitive processes that support relational reasoning, such as relational thinking (Alexander, 2016; Halford et al., 2010) and cognitive control (i.e., working memory, selective attention; Andrews, Birney, & Halford, 2006; Duncan, Chylinski, Mitchell, & Bhandari, 2017). Little work has directly compared inductive and deductive reasoning activations (e.g., Goel & Dolan, 2004), but the evidence to date suggests an overlap in the brain regions that support relational reasoning in both domains (Krawczyk, 2012; Vendetti & Bunge, 2014) Thus, it possible that transfer between inductive and deductive reasoning could occur.

Among the few studies that have assessed transfer between inductive and deductive reasoning, the results have been mixed. Some studies have reported no transfer between the domains (Klauer, Meiser, & Naumer, 2000), while others have reported positive transfer from inductive reasoning instruction to deductive reasoning (Barkl, Porter, & Ginns, 2012; Chapman & Mudar, 2014; Han et al., 2017; K. J. Klauer, Willmes, & Phye, 2002; Roth-Van Der Werf et al., 2002). However, in these studies, transfer was only observed in a subsample of the study (Roth-Van Der Werf et al., 2002), or the gains in deductive reasoning were not retained to the same degree as gains in induction but instead were only apparent at an immediate posttest (Klauer et al., 2002). This latter result was interpreted as the immediate gains stemming from test-retest effects, and not from transfer. Even though there is some evidence of transfer between inductive and deductive reasoning, the conditions under which it occurs have yet to be determined.

Considering the complexity of reasoning skills, including the involvement of several cognitive processes and metacognitive influences (Alexander, 2016; Halford, Wilson, & Phillips, 1998), we need a deep, mechanistic understanding of the cognitive processes that overlap between tasks to predict transfer. Furthermore, a richer understanding of how improvements may influence problem-solving in the transfer domain is an important step towards identifying the mechanisms that either limit or promote transfer effects in reasoning.

Here we seek to provide such insights. We expand on our prior work to assess how improvements in relational thinking after preparing for the Logic Games section of the LSAT may influence performance on an unpracticed test of rule inference. This test, termed "SET" (based on a commercially available card game, and adapted from Eckstein et al., under review), required participants to infer the rules that relate four colored shapes to one another. In this study, we assessed evidence of transfer to behavior, using response times (RTs) and d-prime. We also measured phasic pupillary dilation patterns from participants as they complete this task. The high temporal resolution of the pupillary measure allowed us test whether completion of the Logic Games course could benefit active rule inference on a laboratory task that bore no overt similarity to the problems assigned in the course.

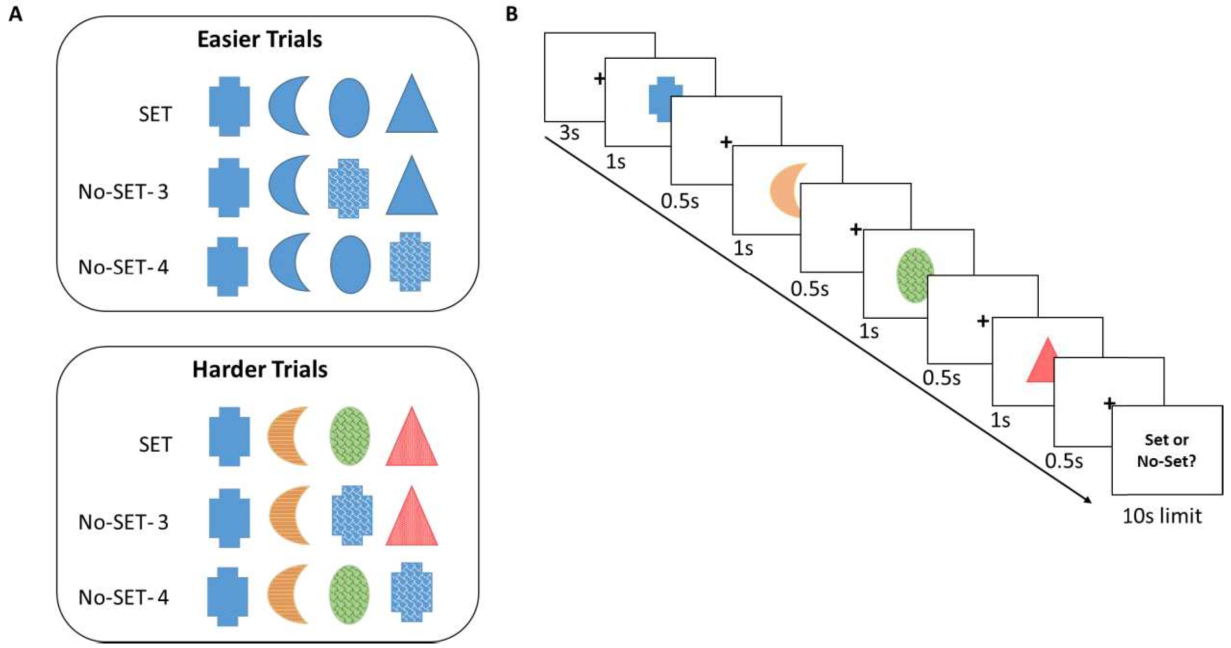


**Figure 1. SET task** (adapted from Eckstein et al., under review). **(A)** Sample stimuli of trials where items spanned in one (top panel) or three dimensions (bottom panel; "harder trials"). Sample 2-span trials are not shown, but were grouped with 1-span trials for analysis (i.e., "easier trials"). Items in all trials spanned in at least one dimension. In half the trials, the items formed a SET (top row of each panel). In the trials where items did not for a SET (i.e., No-SET trials), the rule-incongruent items were presented on the third (middle row of each panel; i.e., No-SET 3 trials) or fourth position (bottom row of each panel; i.e., No-SET 3 trials). **(B)** The trial sequence began after an inter-trial fixation. Each item was presented for 1s and the inter-stimulus fixation for 500ms. Pupillometry analyses were performed over these inter-stimulus fixation windows (see methods for details).

We briefly discuss relevant prior research with the SET task (Eckstein et al., under review), and then present our hypotheses in the next section. The design of the SET task, shown in Figure 1, permits us to probe patterns of pupil dilation as participants accumulated evidence from a series of sequentially presented stimuli, enabling them to generate a hypothesis regarding the governing principle that relates the stimuli to one another. In this task, participants decide whether four items form a "SET," based on whether all four match or all four differ on a given dimension (see Fig 1 for example stimuli and methods for details).

On the trials in which the items form a SET (i.e., SET trials), all four items are necessary to make a final decision (Figure 1A). On trials in which the items do not form a SET (i.e., No-SET trials), a rule-incongruent item appears on either the third position (No-SET-3) or the fourth position (No-SET-4; Fig 1A). These items provide conclusive evidence that the group of items does not form a SET. Thus, on trials where the rule-incongruent item appears third in the sequence of four stimuli (No-SET-3 condition), it is possible to know that the items do not form a SET before seeing the final item. This task design permits us to test the timing of when inductive processes occur. Specifically, to index the temporal dynamics of how people integrate evidence as they generate and test rules, we examined phasic pupil dilations that vary as a function of task demands.

It has been well-established that these phasic pupillary responses occur as norepinephrine release in the brain boosts arousal and dilates the pupils (Eckstein, Guerra-Carrillo, Singley, & Bunge, 2016; Joshi, Li, Kalwani, & Gold, 2017). There are two key pupillary properties of interest that allow us to test our hypothesis in the current study.

The first pupillary property of interest relates to the fact that surprising events elicit phasic pupillary dilations (i.e., the "violation of expectations" response) (Aston-Jones, Rajkowski, Kubiak, & Alexinsky, 1994; Book, Stevens, Pearlson, & Kiehl, 2008; Wetzel, Buttelmann, Schieler, & Widmann, 2016). Based on the terminology in the literature, we will refer to this metric hereafter as the "violation-of-expectation" (VOE) response. This effect has been extensively reported in the context of classic perceptual oddball paradigms, whereby perceptual deviants elicit phasic dilations (for a review see Eckstein et al., 2016). Our prior work with the SET task has shown that this phasic pupillary dilation also occurs in response to conceptual deviants, wherein rule-incongruent items elicit a VOE response relative to rule-congruent items (Eckstein et al., under review). This finding is consistent with related work showing that rule violations are salient and lead to prediction errors (Aston-Jones & Cohen, 2005). Thus, identification of a VOE response evoked by a rule-incongruent item allows us to infer that a piece of evidence was inconsistent with the participant's hypothesized rule – and, by extension, that the participant was engaged in active rule inference while encoding the stimuli.

The second property of interest relates to pupil constrictions or attenuated dilations that occur in response to goal-irrelevant information (Aston-Jones & Cohen, 2005; Gilzenrat et al., 2010). In our prior findings (Eckstein et al., under review), pupil dilation was weaker in response to a rule-congruent item presented immediately *after* a rule-incongruent item (i.e., the fourth item on a trial in which the third item violated the rules – i.e., a No-SET-3 trial) than after a rule-congruent item presented in the same ordinal position (i.e., the fourth item on a trial in which all items were consistent with the rule – i.e., a SET trial). This pattern suggested a greater level of engagement with evidence deemed informative, and that is yet to be integrated with old evidence. Thus, pupillary constrictions following a rule-incongruent item represent a second opportunity to assess whether participants have inferred the relevant rules, and that they have combined their knowledge about rules with knowledge about already-seen items to determine the relevance of incoming evidence.

Thus, by measuring pupil dilation associated with – and following – a rule violation, we can assess rule induction in real time. Building on our prior work, in which we showed improved relational thinking on a transitive inference task after practicing reasoning the Logic Games section of the LSAT (Guerra-Carrillo & Bunge, under review), we hypothesized that improved relational thinking could increase efficiency of the evaluation and integration of evidence, and thus support rule induction. As such, if the Reasoning group were to show an enhanced VOE response to rule-incongruent items from pre- to post-test, we would infer that practicing reasoning led to an increase in active hypothesis generation. If the VOE to a rule-incongruent item in the third position were to show an increased amplitude after reasoning practice, we would infer that they became more effective at evaluating the evidence and inferring rules with little information. Similarly, if we were to observe a pupillary constriction in response to items that appear after a rule-incongruent item, we would infer improvements in the ability to effectively integrate evidence to make predictions about the importance of future evidence. Thus, pupillary responses to both the third and fourth items in the sequence allowed us to test for improvements in rule induction at different points during problem-solving.

In summary, we assessed evidence of transfer from a course that predominantly emphasizes deductive reasoning to performance and the cognitive mechanisms engaged during rule induction. As detailed below, the design of the SET task allowed us to evaluate our predictions at varying gradations of cognitive complexity. We evaluated whether any changes in performance and pupillometry were greater in the Reasoning than in the Comprehension group, to assess whether they could be attributed to an effect of the reasoning course, or rather to learning as a result of repeated exposure to the task.

## Methods

We provide a summary of general methods that we have reported elsewhere (Eckstein et al., under review; Guerra-Carrillo & Bunge, under review), and devote greater attention to aspects of the methods most relevant to the current investigation.

### Ethics statement

The research was approved by the Committee for the Protection of Human Subjects at the University of California, Berkeley. Written informed consent was obtained from all participants.

### Participants and eligibility

We recruited college students planning to take the LSAT within a year. We assigned participants pseudo-randomly to study for one of two sections of the LSAT: Logic Games or Reading Comprehension. The groups matched on age, gender, reasoning, working memory, and LSAT performance at pre-test (Guerra-Carrillo & Bunge, under review). Inclusion criteria included being native English speakers, at least 18 years old, normal/corrected vision, and no history of psychiatric disorders, learning disabilities, or prior LSAT experience.

Forty-nine participants completed the pre- and post-tests and the LSAT course. We excluded two participants because they failed to study for their assigned course. The final sample who prepared for the Logic Games section of the LSAT included 23 students (14 Females, mean age 21.55). The final sample who prepared for the Reading Comprehension section of the LSAT included 24 students (13 females, mean age 21.88). Furthermore, we excluded four participants (2 from Reasoning group) due to technical issues during their eyetracking session, and two participants from the Comprehension group for having more than 60% of trials missing valid pupillometry data and performance below chance levels (where chance was defined as d-prime below 0.51, which is equivalent to scoring below chance in either SET or No-SET trials).

### Summary of procedures

Before and after studying for the LSAT courses, participants completed a battery of nine online cognitive assessments (Hampshire et al., 2012), followed by an in-person testing session.

During the lab sessions, we recorded pupillometry data from participants while they completed the SET task. Participants also completed two other eyetracking tasks during this session: transitive inference (Guerra-Carrillo & Bunge, under review) and a matrix reasoning task. Data from the SET task is the subject of the current investigation. After finishing the eyetracking tasks, participants completed pencil and paper tests, including a standardized test of deductive reasoning, termed Analysis Synthesis (Woodcock et al., 2001), LSAT sample problems, and a survey. The order of tests was the same at both timepoints. Participants were blind to their LSAT group at pre-test, and the experimenters carrying out the testing sessions were blind to the group assignment at both timepoints.

### LSAT courses

Participants studied for either the Logic Games or Reading Comprehension section of the LSAT with a commercially available online course (Kaplan, Inc.) for six weeks. During the intervention, we gave students access only to the instructional material related to their assigned section. Furthermore, we asked participants to space their practice (i.e., study every other day, three times per week), as this has been shown effective in promoting transfer effects (Wang et al., 2014). Participants reported having complied with these instructions, and both groups reported similar studying times (Guerra-Carrillo & Bunge, under review).

The Logic Games section involves solving word problems that contain many rules that must be integrated to find the correct answer (sample problems: https://www.lsac.org/jd/lsat/prep/analytical-reasoning). In the preparatory course, participants learn strategies, such as organizing relational information into sketches, to facilitate deduction and correct rule application. The Reading Comprehension section involves reading long passages and answering multiple-choice questions based on relevant information in the passages (sample problems: https://www.lsac.org/jd/lsat/prep/reading-comprehension). In the course, participants learn strategic reading techniques, such as finding keywords and annotating key ideas. Both courses contained the same number of lessons, used the same instructional format, and emphasized improvement in timing and increased mastery with different question types for the relevant section. Overall study time was approximately 24 hours.

### Eyetracking apparatus and procedures

We recorded binocular pupil data from participants completing the SET task using Tobii T120 Eye Tracker (17-inch monitor, 1280 ×1024 pixel resolution). We sampled at a temporal resolution of 120Hz, with participants sitting at 60 cm from the eyetracker camera. We used E-prime (Psychology Software Tools, Pittsburgh, PA) to present the task stimuli and the Tobii Extension for E-prime to synchronize the timing of the stimulus presentation and ocular events.

We took several precautions to collect high-quality ocular data following recommendations from (Holmqvist et al., 2011), including testing in a quiet room with no windows and controlled luminance Furthermore, participants reported that they did not suffer from medical conditions or used medication that could affect ocular behaviors.

### The SET task

We adapted the task from (Eckstein et al., under review). In this task, participants compare four items and judge whether they form a "SET," based on the symmetry between *all* items on *each* of these dimensions: color, shape, and filling. Four items form a SET if they fulfill either of two symmetry rules on each dimension: 1) all items match on the dimension, or 2) all items mismatch on the dimension. We will refer to these as the "match" or "span" rule respectively. Items in every trial span in at least one dimension, but can span up to three dimensions. This manipulation was designed to increase the cognitive load of the task (see Eckstein et al., under review, for an in-depth discussion about the effects of the span manipulation). Pre-test performance (RT and accuracy) on the lower span levels did not differ in our main condition of interest (No-SET trials), which is a result we replicated from our previous study (Eckstein et al., under review). Thus, for our purposes, we will consider trials where items span in one or two dimensions as "easier" trials, and trials where items span all three dimensions as "harder" trials (see Figure 1 for sample stimuli).

At both time points, participants completed 60 problems, in two blocks of 30 trials. Half of the problems had items that formed a valid SET (i.e., 30 SET trials), and the other half of trials did not form a valid SET (i.e., 30 No-SET trials). The rule-incongruent item in the No-SET trials appeared either in the fourth or third position in an equal number of trials. The first two items provided the foundation for rule induction and hypothesis generation for all trials, but the third or fourth item provided conclusive evidence. The trials were divided equally among the three-span conditions.

Each trial started with a 3-second fixation period. The items then appeared sequentially in the center of the screen, for 1 second each and interleaved by fixation periods of 500ms. After showing all four items, a response prompt ("SET or No-SET?") appeared. This prompt disappeared after participants selected their answer via a button press or 10 seconds had elapsed (see Fig 1B for trial sequence). Stimuli were matched in luminance and varied across trials, so as not to create an expectation that a certain shape will always be associated with a particular rule or condition.

### Behavioral metrics

The behavioral metrics in the task include response times (RTs) and d-prime, which allow us to characterize accuracy for SET and No-SET trials. We bounded the range of hits and false alarm rates to be between 99% and 0.1%, to be able to include participants in our analyses who completed the task perfectly. Also, we used in the logarithmic form of RTs, as they were positively skewed at both timepoints. We performed these corrections separately for each session.

### Preprocessing and analysis of pupil dilation data

We first preprocessed the data, on a trial and subject basis at each timepoint, following similar procedures that we have described elsewhere (Eckstein et al., under review). Briefly, we averaged pupil diameter values from both eyes, then removed artifacts and interpolated the data using a local loess regression model (Cleveland, Grosse, & Shyu, 1992). This latter step removed data points that fell more than five standard deviations from the local mean, based on 160 consecutive time points (~1,333 ms), and interpolated small gaps of missing data within a window of 50 data points (~416ms). We then downsampled and smoothed the data to 20hz using a 100ms rolling average and 3-point smoother. This procedure resulted in a local average of pupil diameter for every 50ms.

We then calculated pupil dilation values relative to a trial-level baseline, calculated as the average of the pupil diameter across the first 200ms of the trial. This approach allowed us to assess pupillary dilations during the task relative to a normalized baseline for each trial and subject. Finally, we used these normalized dilation data to calculate event-related changes in pupillary dilation in response to an item, which we quantified as the difference between pupil dilation associated with the current versus the previous item during the 500ms inter-stimulus fixation proceeding the item of interest (see Fig 1B). Quantifying relative changes in pupil dilation during that fixation windows accommodates for the natural pupillary time courses (Loewenfeld & Lowenstein, 1993), and permits us to directly compare dilation patterns across items and conditions when the visual stimulus was identical. We used these final values to compute the violation-of-expectation (VOE) pupillary dilation and the constrictions following the rule-incongruent item.

We indexed the pupillary VOE response as the difference in mean amplitude of the pupil dilation evoked by the rule-incongruent item (i.e., the third or fourth item in No-SET-3 and No-SET-4 trials, respectively) vs. the mean amplitude of the pupillary response evoked by a rule-congruent item in the same ordinal position in a SET trial (i.e., the third or fourth item in the SET trials). We indexed the constrictions following the rule-incongruent items as the difference between the mean amplitude of the pupillary response evoked by the rule-incongruent item and the pupillary response to the item that followed it (i.e., the fourth vs. third item in No-SET-3 trials).

We acknowledge that changes in latency of the VOE response from pre- to post-test have the potential to bias the magnitude of the VOE response, given that we measure the response in the 500ms time window following the presentation of the item of interest. However, considering the normative time course of pupillary dilations, whereby dilations peak between 1-1.5s after the onset of a stimulus (Loewenfeld & Lowenstein, 1993), our VOE metric should still capture the majority of the pupillary response and the peak dilation amplitude elicited by the items.

### Statistical Analysis

We quantified evidence in support of our hypotheses with Bayesian tests because they provide more explanatory power in situations in which a null effect could be informative (i.e., lack of transfer). Thus, we used Bayes Factor ($BF_{10}$) to assess how likely our data are to support the model testing the hypothesized effect of transfer compared the null model (i.e., no transfer). Thus, we can also quantify the strength for the null hypothesis (Wagenmakers et al., 2017). That is the ratio of the marginal likelihood of the two hypotheses. For all analyses, we used a uniform distribution of prior probabilities with default Cauchy prior scales from the BayesFactor R package (Morey et al., 2015).

We used Bayesian t-tests to evaluate the level of support for the prediction that the Reasoning group would improve on the behavioral and pupillary metrics. If there was evidence that the means within each group differed between timepoints, we used Bayesian mixed regressions to assess the probability that those changes could be attributed to reasoning practice beyond test-retest alone or subject variance. Thus, we quantified the strength of evidence in favor of including the Group *x* Time term relative to a model containing both main effects. We modeled subject variance as a random nuisance factor, but the model design was otherwise equivalent to a 2x2 repeated-measures ANOVA. We conducted separate models for easier and harder span trials. Considering our interest in model comparison, we report $BF_{10}$, and interpret this metric in accordance with prior work (Wagenmakers et al., 2017): $BF_{10} > 1$: data provide positive evidence for the hypothesis, $BF_{10} > 3$: moderate evidence, $BF_{10} > 10$: strong evidence. The inverse applies for the null hypothesis ($1/BF_{10}$). We performed all analyses in Python 3.6 and R 3.2.

## Results

### Probing behavioral scores for evidence of transfer

Participants had high accuracy and fast response times (RTs) both at pre-test and post-test (Fig 2). There was moderate evidence against our hypothesis that the Reasoning group would improve in d-prime scores and RTs between timepoints on either the easier or harder problems (see Table 1 for this and subsequent statistics). There was also moderate evidence that the active control (Comprehension) group's d-prime scores did not change between timepoints, but there was moderate evidence that the Comprehension group became faster at accurately solving the easier SET trials ($\Delta$ RTs $\approx$ 23%) and No-SET trials ($\Delta$ RTs $\approx$ 25%) from pre- to post-test. However, there was no evidence indicating that this reduction in RTs could be attributed to having prepared for the Reading Comprehension course during the intervention. Indeed, the data provided the strongest support for the model that only included the fixed effect of Time for both easier SET ($BF_{10} \approx 9.10 \pm 1.09$) and No-SET ($BF_{10} \approx 26.11 \pm 1.68$) trials. By contrast, adding the Group x Time term worsened the performance of both models (SET $BF_{10} \approx 1.71 \pm 1.93$; No-SET $BF_{10} \approx 5.03 \pm 1.54$). Thus, changes in RTs in the Comprehension group were likely due to individual differences and test-retest effects. In sum, there was moderate evidence indicating a lack of transfer from practicing reasoning with the Logic Games section of the LSAT to performance on the SET task. However, performance from both groups was already very fast and accurate at pre-test.

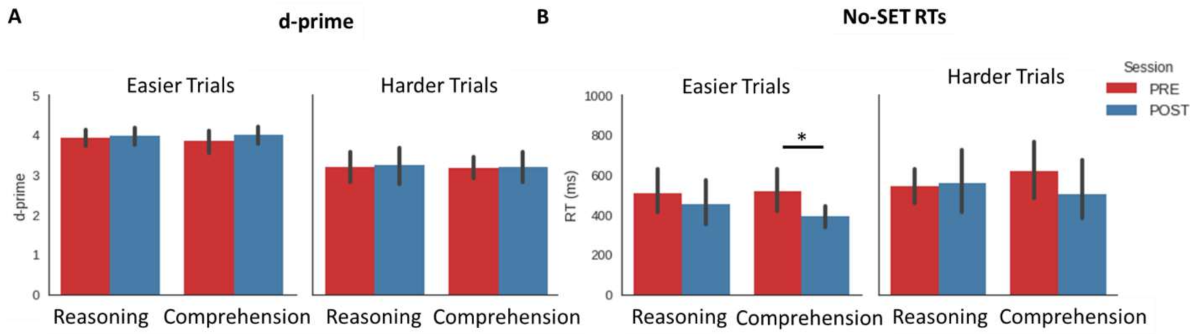


**Figure 2. Behavioral outcomes. (A)** Mean d-prime scores for easier and harder trials for each group at each timepoint. The d-prime scores were calculated with accuracy on SET and No-SET trials. **(B)** Mean response times (RTs) in milliseconds for No-SET, easier and harder trials for each group at each timepoint. RT data for SET trials presented in Table 1. Analysis performed on the log-transformed data. Error bars are 95% C.I estimated with 5000 bootstrap iterations. *Moderate evidence that the Comprehension group showed greater improvements across timepoints. See Table 1 for detailed statistics.

**Table 1.** Evidence of transfer to behavioral metrics

| | | Evidence supporting $H_1$<br>$H_{1\,=}$ change between sessions (POST != PRE)<br>$BF_{10} \approx P(H_1 \mid \text{data}) / P(H_0 \mid \text{data})$ | | | | | |
|---|---|---|---|---|---|---|---|
| | | Reasoning | | | Comprehension | | |
| Metric | Trial Type | PRE<br>*M [95%CI]* | POST<br>*M [95%CI]* | *$BF_{10}$*<br>*(%error)* | PRE<br>*M [95%CI]* | POST<br>*M [95%CI]* | *$BF_{10}$*<br>*(%error)* |
| d-prime | Easier | 3.94<br>[ 3.92, 3.95] | 3.98<br>[ 3.97, 4.00] | 0.23$^{\bullet}$<br>(±0.03) | 3.86<br>[3.84, 3.88] | 4.00<br>[ 3.99, 4.01] | 0.30$^{\bullet}$<br>(±0.01) |
| | Harder | 3.18<br>[ 3.15, 3.20] | 3.23<br>[ 3.20, 3.26] | 0.23$^{\bullet}$<br>(±0.03) | 3.19<br>[3.17, 3.21] | 3.20<br>[ 3.17, 3.22] | 0.24$^{\bullet}$<br>(±0.03) |
| RTs No-SET (log) | Easier | 6.12<br>[ 6.11, 6.13] | 5.98<br>[ 5.96, 5.99] | 1.03<br>(±<0.00) | 6.16<br>[6.15, 6.17] | 5.90<br>[ 5.89, 5.91] | 7.43*<br>(±<0.00) |
| | Harder | 6.25<br>[ 6.23, 6.26] | 6.23<br>[ 6.22, 6.25] | 0.23$^{\bullet}$<br>(±0.03) | 6.26<br>[6.25, 6.28] | 6.07<br>[ 6.06, 6.08] | 0.74<br>(±<0.00) |
| RTs SET (log) | Easier | 6.30<br>[ 6.29, 6.32] | 6.17<br>[ 6.16 6.18] | 0.63<br>(±0.01) | 6.25<br>[6.24, 6.26] | 6.03<br>[ 6.02, 6.04] | 6.46*<br>(±<0.00) |
| | Harder | 6.50<br>[ 6.48, 6.52] | 6.42<br>[ 6.40, 6.43] | 0.32$^{\bullet}$<br>(±0.03) | 6.5<br>[ 6.50, 6.52] | 6.29<br>[ 6.28, 6.30] | 1.11<br>(±<0.00) |

*M[95%CI]: Mean 95% confidence intervals,* calculated with 5000 bootstrap iterations. $H_1$ assessed with Bayesian paired t-tests. No evidence for Group *x* Time interactions. Estimations made using BayesFactor's (Morey et al., 2015) default Cauchy prior scale $r = \frac{1}{2}\sqrt{2}$ and prior uniform probability to the models. Approximate classification scheme to interpret Bayes factors from (Wagenmakers et al., 2017):

*Moderate evidence for $H_1$,, $^{\bullet}$ Moderate evidence for $H_0$.

### Probing pupil dilation patterns for evidence of transfer

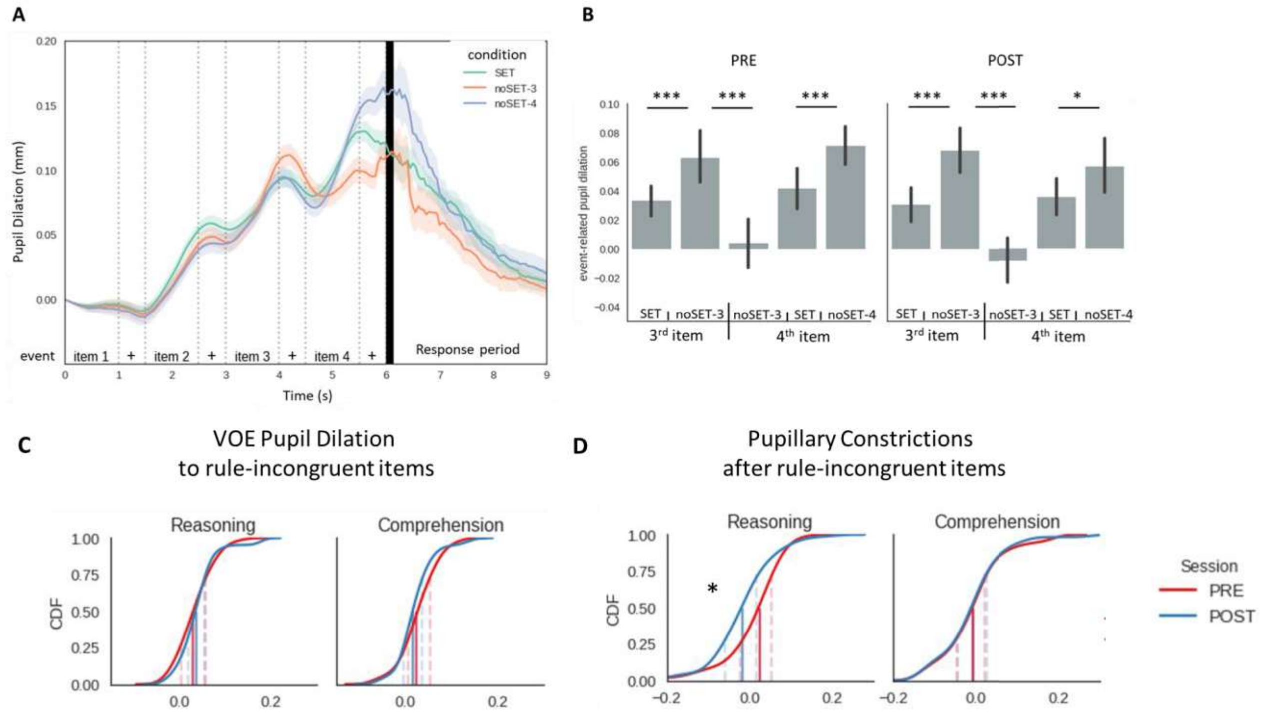


**Figure 3. Pupillary outcome metrics and evidence of transfer. (A)** Pupillary timecourse collapsed across groups and testing sessions. The lines indicate the time course of pupil dilations on SET (green), and No-SET trials, where the rule-incongruent item was presented in the third (orange; No-SET 3) or fourth position (purple; No-SET 4). The ribbons are 95% C.I estimated with 5000 bootstrap iterations. The x-axis denotes the time in the trial from the onset of the first item to the response period. Vertical dotted lines after the third and fourth item, indicate the time windows where we probed the pupil responses of interest. **(B)** Event-related pupil responses to the third or fourth item on SET, No-SET 3, and No-SET 4 trials. We calculated the VOE responses as the difference between the pupil dilation elicited by a rule-incongruent item (i.e., in No-SET trials) with a rule-congruent item (i.e., in SET trials) appearing in the same position. The fourth item in No-SET 3 trials reflects pupillary constrictions following the rule-incongruent item. Error bars are 95% C.I estimated with 5000 bootstrap iterations **(C)** Cumulative distribution functions (CDF) for VOE and **(D)** pupillary constrictions following the rule-incongruent item for each group at each timepoint. Vertical lines denote medians (solid), and 25th or 75th percentiles (dotted lines left or right to the median, respectively). Classification scheme to interpret Bayes factors ($BF_{10}$) used to quantify the strength of evidence in support of the model testing changes between timepoints: ***Extreme, *Moderate. See Table 2 for detailed statistics.

Before testing for the effects of transfer, we examined whether rule-incongruent items elicited a VOE response at either timepoint. Replicating our prior findings (Eckstein et al., under review) with the pretest data and across both groups, we found strong evidence that rule-incongruent items on No-SET trials elicited a larger pupil dilation than items presented on the same ordinal position on SET trials (Fig 3A - 3B). Indeed, a VOE response occurred on both No-SET-3 trials ($BF_{10} \approx$ 121.85 ±<0.00; $\text{VOE}_M$ = 0.029, 95% CI = [0.028, 0.030]) and No-SET-4 trials ($BF_{10} \approx$ 93.16 ±<0.00; $\text{VOE}_M$ =0.028, 95% CI = [0.027, 0.030]). A similar pattern of results emerged at post-test (Fig 3B), such that both groups had a VOE response on No-SET-3 trials ($BF_{10} \approx$ 35588.07 ±<0.00; $\text{VOE}_M$ =0.037, 95% CI = [0.036, 0.038]) and No-SET-4 trials ($BF_{10} \approx$ 5.03 ±<0.00; $\text{VOE}_M$ =0.021, 95% CI = [0.020, 0.023]). Thus, there is evidence that the groups engaged in active rule induction, even when – in the case of No-SET-3 trials – they only had two pieces of evidence as a foundation to start formulating their hypotheses.

Having found strong evidence that a VOE response occurred at both timepoints, we sought to test for effects of reasoning training. We found moderate evidence against our hypothesis that the Reasoning group would have a larger VOE response to rule-incongruent items from pre- to post-test (Fig 3C; for this and subsequent statistics, see Table 2). There was also moderate evidence that the VOE responses of the Comprehension group did not change between timepoints. These results were similar for trials where the rule-incongruent item appeared third or fourth. Thus, these results suggest that there was no effect of reasoning instruction and no change between timepoints for either group in the VOE response.

**Table 2.** Evidence of transfer in pupillary responses indexing rule-induction

| | Evidence supporting $H_1$<br>$H_1$ = change between sessions (POST != PRE)<br>$BF_{10} \approx P(H_1 \mid \text{data}) / P(H_0 \mid \text{data})$ | | | | | |
|---|---|---|---|---|---|---|
| | Reasoning | | | Comprehension | | |
| Pupillary Response | PRE<br>*M [95%CI]* | POST<br>*M [95%CI]* | $BF_{10}$<br>*(%error)* | PRE<br>*M [95%CI]* | POST<br>*M [95%CI]* | $BF_{10}$<br>*(%error)* |
| VOE (No-SET 3) | 0.034<br>[0.033, 0.035] | 0.045<br>[0.044, 0.046] | 0.34•<br>(±0.02) | 0.025<br>[0.024, 0.026] | 0.029<br>[0.028, 0.030] | 0.26•<br>(±0.02) |
| VOE (No-SET 4) | 0.030<br>[0.029, 0.032] | 0.033<br>[0.032, 0.034] | 0.23•<br>(±0.03) | 0.027<br>[0.025, 0.028] | 0.010<br>[0.009, 0.012] | 0.54<br>(±0.01) |
| Pupil Constriction following rule-incongruent item | 0.011<br>[0.009, 0.013] | -0.010<br>[-0.012, -0.008] | 3.94*<br>(±0.01) | -0.005<br>[-0.007, -0.004] | -0.007<br>[-0.009, -0.006] | 0.24•<br>(±0.02) |

*M [95%CI]: Mean with 95% confidence intervals,* calculated with 5000 bootstrap iterations. $H_1$ assessed with Bayesian paired t-tests. No evidence for Group x Time interactions. Estimations made using BayesFactor's (Morey et al., 2015) default Cauchy prior scale $r = \frac{1}{2}\sqrt{2}$ and prior uniform probability to the models. Approximate classification scheme to interpret Bayes factors from (Wagenmakers et al., 2017):

*Moderate evidence for $H_1$,, • Moderate evidence for $H_0$.

Next, we examined evidence of changes in rule induction at a different point during problem-solving, by probing pupillary constrictions in response to the item immediately following the rule-incongruent item. We found moderate evidence that the Reasoning group had larger constrictions following the rule-incongruent item from pre- to post-test (Fig 3D, for this and subsequent statistics, see Table 2). By contrast, there was moderate evidence that the Comprehension group did not change on this metric between timepoints. We examined whether changes in the Reasoning group could be attributed to the intervention. The data provided no evidence either in support of or against the Group x Time interaction term ($BF_{10} \approx 0.45 \pm 1.46$). There was weak support for the effect of Time ($BF_{10} \approx 1.25 \pm 9.73$). Thus, there is no evidence of transfer from practicing reasoning with the Logic Games section of the LSAT to the way people engage in rule induction on the SET task, but rather, practice-related improvements in the Reasoning group (i.e., test-retest effects) or possibly regression to the mean.

## Discussion

We sought to examine transfer from practicing reasoning in a context that primarily involves deductive reasoning to performance on a rule induction task termed SET (Eckstein et al., 2018). We probed pupillary dilation patterns to characterize how practicing reasoning with a 6-week preparatory course for the Logic Games section of the LSAT exam may influence how rule induction processes unfold during problem-solving.

We found moderate evidence against the hypothesized transfer effect on both behavioral and pupillometry outcome measures, which included changes in RTs, d-prime, and VOE response to rule-incongruent items. However, we found moderate evidence of practice-related (i.e., test-retest) improvements in the pupil dilation patterns indexing changes in the way the Reasoning group engaged in rule induction. Indeed, the group showed greater pupillary constrictions to pieces of evidence that provided redundant information and appeared after items that provided conclusive evidence about the rule that applied to the trial. The active control group that prepared for the Reading Comprehension section of the exam did not show those changes. Despite the changes observed in the Reasoning group on this ocular index of active rule induction, there was no evidence that the improvements were directly related to the intervention.

We had expected to find broader evidence of transfer from practicing reasoning with the Logic Games course to rule-inductive processes. Indeed, we had previously shown that the Reasoning group improved on a composite measure of four reasoning tests, which combined measures of inductive and deductive reasoning, as well as in relational thinking efficiency, as indexed with eye gaze data collected while they solved a transitive inference task that taps into deductive reasoning (Guerra-Carrillo & Bunge, under review). These improvements were larger than those of the Comprehension group. The behavioral changes in the Reasoning group were consistent with prior work from our lab with a sample of young adults who prepared for all the sections of the LSAT with a three-month course (Mackey et al., 2015). The sample who prepared for the LSAT also showed changes in structural and functional connectivity between key nodes in the frontoparietal network (Mackey et al., 2013, 2012), which have been implicated in relational reasoning (Krawczyk, 2012; Prado et al., 2011). Considering all these prior findings and the fact that inductive and deductive reasoning share many of the same underlying cognitive and neural mechanisms, we had expected to find evidence of positive transfer in the present study that could be directly attributed to the learning experience. However, our present findings support results we had previously considered inconclusive in regards to transfer from studying for all sections of the LSAT to performance on a rule-generation task (Mackey et al., 2015).

It is possible that lack of transfer we observe in our current study stems from a large degree of individual differences, for example in the way participants represent rules (Eckstein et al., under review). The current task could also be too easy and not sensitive enough to detect transfer effects, at least at a behavioral level. However, it is also possible that the lack of transfer in our study reflects that improving in relational thinking may not be sufficient to boost other processes engaged in rule induction, such as generating hypotheses with minimal evidence about rules that relate the observed evidence. This latter interpretation could explain why there was evidence that the Reasoning group had bigger increases in pupillary constrictions following a rule-incongruent item (i.e., fourth item on No-SET trials where the incongruent item was presented third), but not larger VOE response to the rule-incongruent item itself at post-test (i.e., rule-incongruent item appearing on the third position).

There are at least four limitations of our study to consider. First, we could not explain the Comprehension's group reduction of RTs on easier trials with any of the pupillary responses we examined. It is plausible that other changes in pupillary responses that we were not able to examine with the version of the task we used in the present study (e.g., metrics of rule representation, see Eckstein et al., under review) could have influenced the changes in RTs in this group.  Second, the sample size is relatively small. As we have discussed previously (Guerra-Carrillo & Bunge, under review), recruitment and retention were an issue considering that we needed participants willing to study for only one section of the LSAT, which was a problem for students seeking to take the exam immediately. Third, there seemed to be more noise in the pupillometry data at post-test during the response period. Although the noise occurred in the response period, it is possible that we may not be fully capturing the entire pupillary time course in response to the final items presented during a trial. In our previous study with the SET task (Eckstein et al., under review), we had found that the responses elicited by the rule-incongruent item presented in the last position extended into the first second of the response period. A related issue is that changes in rule induction could also result in shifts in latency of the pupillary responses we used as an indicator of active rule inference. If so, the period where we calculated our pupillary metrics may not have included the entire pupillary response. Even though these two issues have the potential to introduce some noise in our pupillary metrics, we likely still captured all or the majority of the pupillary responses at both pre- and post-test considering the natural time course of pupil dilations. Indeed, we measured dilation amplitude during a time window that could contain the peak dilation response.

Despite these limitations, our null results can inform the growing literature that has aimed to delineate the extent of transfer resulting from cognitive learning experiences (Noack et al., 2014) in the following ways. First, we quantified the strength of the null hypothesis of no transfer, and our approach could directly inform how priors are set in future work and contribute to a growing body of evidence about transfer effects (Wagenmakers et al., 2017). Compared to frequentist methods, this approach is also more powerful for testing hypotheses about transfer with smaller sample sizes, wherein several measures are necessary to delineate boundaries of transfer between domains (Lövdén et al., 2010).  Second, we obtained evidence from pupil dilation patterns that permit us to characterize how rule induction processes unfold during problem-solving. Based on the pupil dilation patterns, we were able to identify a subtle change in the way the Reasoning group actively engaged in rule inference, despite lack of changes in behavior.

Our study illustrates the use of a method that could be employed to obtain insights about the malleability of higher-level cognition in different domains. We hypothesize that these types of ocular metrics could provide sensitive measures of ongoing changes over the course of an intervention that either accompany or precede changes in behavior. This argument has been made for neuroimaging methods (Blakemore & Bunge, 2012), but translates well to eyetracking measures given the strong link between ocular behaviors and cognitive processing (Eckstein et al., 2016; Kowler, 2011).

## General discussion

Over the years, we acquire knowledge, develop new skills, and engage in numerous activities that are rich in novelty and cognitive challenge. How do all these experiences shape us?

For one, there are changes in how the brain networks engage after learning. Spontaneous fluctuations of BOLD activity between regions of a network that support the learned task become more tightly synchronized at rest. These changes give us clues about the history of co-activation that might have repeatedly occurred between those regions during learning, as discussed in Chapter 1. Since publishing that review, there have been other studies using rsfMRI correlations to study brain plasticity resulting from motor rehabilitation (Rjosk et al., 2017), gained skill-based expertise (Cantou et al., 2017), and seeking to understand the role of environmental input in the ontogeny of cortical and subcortical circuits in the human brain (Gabard-Durnam et al., 2016). A better understanding of how networks change with experience can continue to open up opportunities to develop effective rehabilitation therapies, and further our understanding of the plasticity of networks underlying higher-level cognition.

There is also evidence that the long-lasting experience of formal education influences cognitive performance across late adolescence and adulthood in unique ways. In Chapter 2, I showed that educational attainment is more strongly associated with how well people perform on reasoning tests than on other measures, like processing speed, which is more influenced by the age of the person. Also, it turns out that educational experiences modulate previously identified ages of when cognitive functioning is supposedly at its peak. We also provided evidence that although someone's educational history influences their starting level in a new learning context, it has little bearing on how much they can benefit from the new opportunity. An intriguing future direction for this work would be a longitudinal investigation of periods preceding and following critical junctures in formal schooling, such as graduating from high school and attending college, or graduating from college and entering the workforce. More longitudinal work is necessary not only for a scientific perspective but also to provide even stronger arguments for advocates fighting to make higher education more widely accessible.

Finally, I have characterized cognitive mechanisms that support improved performance on reasoning skills following targeted reasoning instruction and practice. One of the mechanisms we identified was improved relational thinking. That is, a group of young adults became more proficient at encoding, maintaining, and integrating visual relations after only a relatively short learning experience that mainly involved dealing with verbal relations and rules. Even in the cases where changes were probably due to having had prior experience with the test, it is still remarkable to observe how cognitive processing adapts and becomes fine-tuned to the demands of a task only after having performed it once. The ability to measure this type of change is something that could be applied to further our understanding of changes that occur from other learning experiences and during development, and may also be a fruitful venue to examine plasticity in late adulthood (Kühn & Lindenberger, 2016).

In sum, we know that people learn and we now know a little more about how that happens.

# References

Agostino, R., Bologna, M., Dinapoli, L., Gregori, B., Fabbrini, G., Accornero, N., & Berardelli, A. (2008). Voluntary, spontaneous, and reflex blinking in Parkinson's disease. *Movement Disorders*, *23*(5), 669–675. http://doi.org/10.1002/mds.21887

Ahern, S. (1978). *Activation and intelligence: pupillometric correlates of individual differences in cognitive abilities*. University of California, Los Angeles.

Ahern, S., & Beatty, J. (1979). Pupillary responses during information processing vary with Scholastic Aptitude Test scores. *Science (New York, N.Y.)*, *205*(4412), 1289–1292.

Al-Abdulmunem, M., & Briggs, S. T. (1999). Spontaneous blink rate of a normal population sample. *International Contact Lens Clinic*, *26*(99), 29–32. http://doi.org/10.1016/S0892-8967(99)00016-4

Albin, R. L., & Mink, J. W. (2006). Recent advances in Tourette syndrome research. *Trends in Neurosciences*, *29*(3), 175–182. http://doi.org/10.1016/j.tins.2006.01.001

Alexander, P. A. (2016). Relational thinking and relational reasoning: harnessing the power of patterning. *Npj Science of Learning*, *1*(1), 16004. http://doi.org/10.1038/npjscilearn.2016.4

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*(1), 20–29. http://doi.org/10.1016/j.jecp.2009.11.003

Alnaes, D., Sneve, M. H., Espeseth, T., Endestad, T., van de Pavert, S. H. P., & Laeng, B. (2014). Pupil size signals mental effort deployed during multiple object tracking and predicts brain activity in the dorsal attention network and the locus coeruleus. *Journal of Vision*, *14*(4), 1–1. http://doi.org/10.1167/14.4.1

Andreassi, J. L. (2000). Pupillary response and behavior. In *Psychophysiology: Human Behavior & Physiological Response* (4th ed., pp. 218–233). Mahwah, N.J: Psychology Press.

Andrews, G., Birney, D., & Halford, G. S. (2006). Relational processing and working memory capacity in comprehension of relative clause sentences. *Memory & Cognition*, *34*(6), 1325–1340. http://doi.org/10.3758/BF03193275

Arnsten, A., & Li, B. (2004). Neurobiology of executive functions: Catecholamine influences on prefrontal cortical functions. *Biological Psychiatry*, *57*, 1277–1384. http://doi.org/10.1016/j.bps.2004.08.019

Asato, M. R., Sweeney, J. A., & Luna, B. (2006). Cognitive processes in the development of TOL performance. *Neuropsychologia*, *44*(12), 2259–2269. http://doi.org/10.1016/j.neuropsychologia.2006.05.010

Aslin, R. N. (2007). What's in a look? *Developmental Science*, *10*(1), 48–53. http://doi.org/10.1111/J.1467-7687.2007.00563.X

Aslin, R. N., & McMurray, B. (2004). Automated corneal-reflection eye tracking in infancy:

Methodological developments and applications to cognition. *Infancy*, *6*(2), 155–163.

Aston-Jones, G., Chiang, C., & Alexinsky, T. (1991). Discharge of noradrenergic locus coeruleus neurons in behaving rats and monkeys suggests a role in vigilance. *Progress in Brain Research*, *88*, 501–520.

Aston-Jones, G., & Cohen, J. D. (2005). Adaptive gain and the role of the locus coeruleus-norepinephrine system in optimal performance. *The Journal of Comparative Neurology*, *493*(1), 99–110. http://doi.org/10.1002/cne.20723

Aston-Jones, G., & Cohen, J. D. (2005). AN INTEGRATIVE THEORY OF LOCUS COERULEUS-NOREPINEPHRINE FUNCTION: Adaptive Gain and Optimal Performance. *Annual Review of Neuroscience*, *28*(1), 403–450. http://doi.org/10.1146/annurev.neuro.28.061604.135709

Aston-Jones, G., Rajkowski, J., & Cohen, J. (1999). Role of locus coeruleus in attention and behavioral flexibility. *Biological Psychiatry*, *46*(9), 1309–1320. http://doi.org/10.1016/S0006-3223(99)00140-7

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. *Psychonomic Bulletin and Review*, *22*(2), 366–377. http://doi.org/10.3758/s13423-014-0699-x

Awh, E., Armstrong, K. M., & Moore, T. (2006). Visual and oculomotor selection: links, causes and implications for spatial attention. *Trends in Cognitive Sciences*, *10*(3), 124–130. http://doi.org/10.1016/j.tics.2006.01.001

Bacher, L. F. (2014). Development and manipulation of spontaneous eye blinking in the first year: relationships to context and positive affect. *Developmental Psychobiology*, *56*, 783–796. http://doi.org/10.1002/dev.21148

Bacher, L., & Smotherman, W. P. (2004). Spontaneous eye blinking in human infants: a review. *Developmental Psychobiology*, *44*(2), 95–102. http://doi.org/10.1002/dev.10162

Baillargeon, R., Spelke, E. S., & Wasserman, S. (1985). Object permanence in five-month-old infants. *Cognition*, *20*(3), 191–208.

Baldassarre, A., Lewis, C. M., Committeri, G., & Snyder, A. Z. (2012). Individual variability in functional connectivity predicts performance of a perceptual task. *PNAS*, *109*(9), 3516–3521. http://doi.org/10.1073/pnas.1113148109/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1113148109

Barbato, G., Ficca, G., Muscettola, G., Fichele, M., Beatrice, M., & Rinaldi, F. (2000). Diurnal variation in spontaneous eye-blink rate. *Psychiatry Research*, *93*(2), 145–151. http://doi.org/10.1016/S0165-1781(00)00108-6

Barkl, S., Porter, A., & Ginns, P. (2012). Cognitive training for children: Effects on inductive reasoning, deductive reasoning, and mathematics achievement in an Australian school setting. *Psychology in the Schools*, *49*(9), 828–842. http://doi.org/10.1002/pits.21638

Barkley-Levenson, E., & Galván, A. (2016). Eye blink rate predicts reward decisions in adolescents. *Developmental Science*, 1–11. http://doi.org/10.1111/desc.12412

Barnes, A., Bullmore, E. T., & Suckling, J. (2009). Endogenous Human Brain Dynamics Recover Slowly Following Cognitive Effort. *PLoS ONE*, *4*(8), 6. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/19680553

Barrow, L., & Malamud, O. (2015). Is College a Worthwhile Investment? *Annual Review of Economics*, *7*, 519–555. http://doi.org/10.1146/annurev-economics-080614-115510

Bavelier, D., Green, C. S., Pouget, A., & Schrater, P. (2012). Brain Plasticity Through the Life Span : Learning to Learn and Action Video Games. *Annual Review of Neuroscience*, *35*, 391–416. http://doi.org/10.1146/annurev-neuro-060909

Bavelier, D., Levi, D. M., Li, R. W., Dan, Y., & Hensch, T. K. (2010). Removing Brakes on Adult Brain Plasticity: From Molecular to Behavioral Interventions. *Journal of Neuroscience*, *30*(45), 14964–14971. http://doi.org/10.1523/JNEUROSCI.4812-10.2010

Bayer, H. M., & Glimcher, P. W. (2005). Midbrain Dopamine Neurons Encode a Quantitative Reward Prediction Error Signal. *Neuron*, *47*(1), 129–141. http://doi.org/10.1016/j.neuron.2005.05.020

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin*, *91*(2), 276.

Beatty, J., & Lucero-Wagoner, B. (2000). The pupillary system. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology* (2nd ed., pp. 142–162). New York, NY, US: Cambridge University Press.

Beaulieu, J.-M., & Gainetdinov, R. R. (2011). The physiology, signaling, and pharmacology of dopamine receptors. *Pharmacological Reviews*, *63*(1), 182–217. http://doi.org/10.1124/pr.110.002642.182

Bediou, B., Adams, D. M., Mayer, R. E., Tipton, E., Green, C. S., & Bavelier, D. (2018). Meta-analysis of action video game impact on perceptual, attentional, and cognitive skills. *Psychological Bulletin*, *144*(1), 77–110. http://doi.org/10.1037/bul0000130

Behzadi, Y., Restom, K., Liau, J., & Liu, T. T. (2007). A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *NeuroImage*, *37*(1), 90–101. http://doi.org/10.1016/j.neuroimage.2007.04.042

Bergman Nutley, S., Söderqvist, S., Bryde, S., Thorell, L. B., Humphreys, K., & Klingberg, T. (2011). Gains in fluid intelligence after training non-verbal reasoning in 4-year-old children: A controlled, randomized study. *Developmental Science*, *14*(3), 591–601. http://doi.org/10.1111/j.1467-7687.2010.01022.x

Berridge, K. C. (2007). The debate over dopamine’s role in reward: the case for incentive salience. *Psychopharmacology*, *191*(3), 391–431. http://doi.org/10.1007/s00213-006-0578-x

Bijleveld, E., Custers, R., & Aarts, H. (2012). Human Reward Pursuit: From Rudimentary to

Higher-Level Functions. *Current Directions in Psychological Science*, *21*(3), 194–199. http://doi.org/10.1177/0963721412438463

Blakemore, S. J., & Bunge, S. A. (2012). At the nexus of neuroscience and education. *Developmental Cognitive Neuroscience*, *2*(SUPPL. 1), 1–5. http://doi.org/10.1016/j.dcn.2012.01.001

Blaser, E., Eglington, L., Carter, A. S., & Kaldy, Z. (2014). Pupillometry Reveals a Mechanism for the Autism Spectrum Disorder (ASD) Advantage in Visual Tasks. *Scientific Reports*, *4*. http://doi.org/10.1038/srep04301

Blundon, J. a, & Zakharenko, S. S. (2013). Presynaptic Gating of Postsynaptic Synaptic Plasticity. *The Neuroscientist*, *19*(5), 465–478. http://doi.org/10.1177/1073858413482983

Bochynska, A., & Laeng, B. (2015). Tracking down the path of memory: eye scanpaths facilitate retrieval of visuospatial information. *Cognitive Processing*, *16*(1), 159–163. http://doi.org/10.1007/s10339-015-0690-0

Boersma, F., Wilton, K., Barham, R., & Muir, W. (1970). Effects of arithmetic problem difficulty on pupillary dilation in normals and educable retardates. *Journal of Experimental Child Psychology*, *9*(2), 142–155. http://doi.org/10.1016/0022-0965(70)90079-2

Bologna, M., Fasano, A., Modugno, N., Fabbrini, G., & Berardelli, A. (2012). Effects of subthalamic nucleus deep brain stimulation and l-dopa on blinking in Parkinson’s disease. *Experimental Neurology*, *235*(1), 265–272. http://doi.org/10.1016/j.expneurol.2012.02.004

Bonaccorsi, J., Berardi, N., & Sale, A. (2014). Treatment of amblyopia in the adult: insights from a new rodent model of visual perceptual learning. *Frontiers in Neural Circuits*, *8*(July), 82. http://doi.org/10.3389/fncir.2014.00082

Book, G., Stevens, M. C., Pearlson, G., & Kiehl, K. A. (2008). Fusion of fMRI and the Pupil Response During an Auditory Oddball Task.

Breedlove, S. M., Watson, N. V., & Rosenzweig, M. R. (2010). *Biological Psychology: An Introduction to Behavioral, Cognitive, and Clinical Neuroscience* (6th ed.). Sunderland, MA: Sinauer Associates, Inc.

Brinch, C. N., & Galloway, T. A. (2012). Schooling in adolescence raises IQ scores. *Proceedings of the National Academy of Sciences*, *109*(2), 425–430. http://doi.org/10.1073/pnas.1106077109

Brod, G., Bunge, S. A., & Shing, Y. L. (2017). Does One Year of Schooling Improve Children’s Cognitive Control and Alter Associated Brain Activation? *Psychological Science*, *28*(7), 967–978. http://doi.org/10.1177/0956797617699838

Buckner, R. L., Krienen, F. M., & Yeo, B. T. T. (2013). Opportunities and limitations of intrinsic functional connectivity MRI. *Nature Neuroscience*, *16*(7), 832–7. http://doi.org/10.1038/nn.3423

Buckner, R. L., & Vincent, J. L. (2007). Unrest at rest: default activity and spontaneous network correlations. *NeuroImage*, *37*(4), 1091-6; discussion 1097-9. http://doi.org/10.1016/j.neuroimage.2007.01.010

Bunge, S. a., & Wright, S. B. (2007). Neurodevelopmental changes in working memory and cognitive control. *Current Opinion in Neurobiology*, *17*, 243–250. http://doi.org/10.1016/j.conb.2007.02.005

Burkhouse, K. L., Siegle, G. J., Woody, M. L., Kudinova, A. Y., & Gibb, B. E. (2015). Pupillary reactivity to sad stimuli as a biomarker of depression risk: Evidence from a prospective study of children. *Journal of Abnormal Psychology*, *124*(3), 498–506. http://doi.org/10.1037/abn0000072

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

Cabestrero, R., Crespo, A., & Quirós, P. (2009). Pupillary dilation as an index of task demands. *Perceptual and Motor Skills*, *109*(3), 664–678. http://doi.org/10.2466/pms.109.3.664-678

Calhoun, V. D., Liu, J., & Adalı, T. (2009). A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data. *NeuroImage*, *45*, 163–172. http://doi.org/10.1016/j.neuroimage.2008.10.057

Calvo, M. G. (2001). Working memory and inferences: Evidence from eye fixations during reading. *Memory*, *9*(4–6), 365–381. http://doi.org/10.1080/09658210143000083

Caplan, R., & Guthrie, D. (1994). Blink rate in childhood schizophrenia spectrum disorder. *Biological Psychiatry*, *35*(4), 228–234.

Cardak, B. A., & Ryan, C. (2009). Participation in higher education in Australia: Equity and access. *Economic Record*, *85*(271), 433–448. http://doi.org/10.1111/j.1475-4932.2009.00570.x

Casey, B. J., Galván, A., & Somerville, L. H. (2015). Beyond simple models of adolescence to an integrated circuit-based account: A commentary. *Developmental Cognitive Neuroscience*, *17*, 128–130. http://doi.org/10.1016/j.dcn.2015.12.006

Cattell, R. B. (1971). *Abilities: Their structure, growth, and action*. Boston: Houghton Mifflin.

Ceci, S. J. (1991). How much does schooling influence general intelligence and its cognitive components? A reassessment of the evidence. *Developmental Psychology*, *27*(5), 703–722. http://doi.org/10.1037/0012-1649.27.5.703

Chai, X. J., Ofen, N., Gabrieli, J. D. E., & Whitfield-Gabrieli, S. (2014). Selective Development of Anticorrelated Networks in the Intrinsic Functional Organization of the Human Brain. *Journal of Cognitive Neuroscience*, *26*(3), 501–513. http://doi.org/10.1162/jocn_a_00517

Chan, R. C. K., & Chen, E. Y. H. (2004). Blink Rate Does Matter: A Study of Blink Rate, Sustained Attention, and Neurological Signs in Schizophrenia. *The Journal of Nervous and Mental Disease*, *192*(11), 781–783. http://doi.org/10.1097/01.nmd.0000144697.48042.eb

Chapman, S. B., & Mudar, R. A. (2014). Enhancement of cognitive and neural functions through complex reasoning training: evidence from normal and clinical populations. *Frontiers in Systems Neuroscience*, *8*(April), 1–7. http://doi.org/10.3389/fnsys.2014.00069

Chatham, C. H., Frank, M. J., & Munakata, Y. (2009). Pupillometric and behavioral markers of a developmental shift in the temporal dynamics of cognitive control. *Proceedings of the National Academy of Sciences*, *106*(14), 5529–5533.

Chen, Z., Honomichl, R., Kennedy, D., & Tan, E. (2016). Aiming to Complete the Matrix: Eye-Movement Analysis of Processing Strategies in Children's Relational Thinking. *Developmental Psychology*, *52*(6), 867–78. http://doi.org/10.1037/dev0000113

Chermahini, S. A., & Hommel, B. (2010). The (b)link between creativity and dopamine: Spontaneous eye blink rates predict and dissociate divergent and convergent thinking. *Cognition*, *115*(3), 458–465. http://doi.org/10.1016/j.cognition.2010.03.007

Chevalier, N., Martis, S. B., Curran, T., & Munakata, Y. (2015). Metacognitive processes in executive control development: the case of reactive and proactive control. *Journal of Cognitive Neuroscience*, *27*(6), 1125–1136. http://doi.org/10.1162/jocn_a_00782

Christoforides, M., Spanoudis, G., & Demetriou, A. (2016). Coping With Logical Fallacies: A Developmental Training Program for Learning to Reason. *Child Development*, *87*(6), 1856–1876. http://doi.org/10.1111/cdev.12557

Cliffordson, C., & Gustafsson, J.-E. (2008). The Effects of age and schooling on intellectual performance: Estimates obtained from analysis of continous variation in age and length of schooling. *Intelligence*, *36*(2), 143–152.

Cole, M. W., Yarkoni, T., Repovs, G., Anticevic, A., & Braver, T. S. (2012). Global Connectivity of Prefrontal Cortex Predicts Cognitive Control and Intelligence. *Journal of Neuroscience*, *32*(26), 8988–8999. http://doi.org/10.1523/JNEUROSCI.0536-12.2012

Colzato, L. S., Van Den Wildenberg, W. P. M., Van Wouwe, N. C., Pannebakker, M. M., & Hommel, B. (2009). Dopamine and inhibitory action control: Evidence from spontaneous eye blink rates. *Experimental Brain Research*, *196*, 467–474. http://doi.org/10.1007/s00221-009-1862-x

Constantinidis, C., & Klingberg, T. (2016). The neuroscience of working memory capacity and training. *Nature Reviews Neuroscience*, *17*(7), 438–449. http://doi.org/10.1038/nrn.2016.43

Cools, R., & D'Esposito, M. (2011). Inverted-U–Shaped Dopamine Actions on Human Working Memory and Cognitive Control. *Biological Psychiatry*, *69*(12), e113–e125. http://doi.org/10.1016/j.biopsych.2011.03.028

Corbetta, M., Akbudak, E., Conturo, T. E., Snyder, A. Z., Ollinger, J. M., Drury, H. A., … Shulman, G. L. (1998). A common network of functional areas for attention and eye movements. *Neuron*, *21*(4), 761–773.

Costa, V. D., & Rudebeck, P. H. (2016). More than Meets the Eye: the Relationship between Pupil Size and Locus Coeruleus Activity. *Neuron*, *89*(1), 8–10.

http://doi.org/10.1016/j.neuron.2015.12.031

Cruz, A. a V, Garcia, D. M., Pinto, C. T., & Cechetti, S. P. (2011). Spontaneous eyeblink activity. *The Ocular Surface*, *9*(1), 29–41. http://doi.org/10.1016/S1542-0124(11)70007-6

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

Damoiseaux, J. S., Rombouts, S. A. R. B., Barkhof, F., Scheltens, P., Stam, C. J., Smith, S. M., & Beckmann, C. F. (2006). Consistent resting-state networks across healthy subjects. *Proceedings of the National Academy of Sciences*, *103*(37), 13848–13853. http://doi.org/10.1073/pnas.0601417103

Dang, J., Xiao, S., Liu, Y., Jiang, Y., & Mao, L. (2016). Individual differences in dopamine level modulate the ego depletion effect. *International Journal of Psychophysiology*, *99*, 121–124. http://doi.org/10.1016/j.ijpsycho.2015.11.013

Daniel, D. B., & Klaczynski, P. A. (2006). Developmental and individual differences in conditional reasoning: Effects of logic instructions and alternative antecedents. *Child Development*, *77*(2), 339–354. http://doi.org/10.1111/j.1467-8624.2006.00874.x

De Havas, J. A., Parimal, S., Soon, C. S., & Chee, M. W. L. (2012). Sleep deprivation reduces default mode network connectivity and anti-correlation during rest and task performance. *NeuroImage*, *59*(2), 1745–1751. http://doi.org/10.1016/j.neuroimage.2011.08.026

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

Demarais, A. M., & Cohen, B. H. (1998). Evidence for image-scanning eye movements during transitive inference. *Biological Psychology*, *49*(3), 229–247. http://doi.org/10.1016/S0301-0511(98)00042-8

den Daas, C., Häfner, M., & de Wit, J. (2013). Out of Sight, Out of Mind: Cognitive States Alter the Focus of Attention. *Experimental Psychology*, *60*(5), 313–323. http://doi.org/10.1027/1618-3169/a000201

Deubel, H., & Schneider, W. X. (1996). Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vision Research*, *36*(12), 1827–1837.

Dewhurst, R., Nyström, M., Jarodzka, H., Foulsham, T., Johansson, R., & Holmqvist, K. (2012). It depends on how you look at it: Scanpath comparison in multiple dimensions with MultiMatch, a vector-based approach. *Behavior Research Methods*, 1079–1100. http://doi.org/10.3758/s13428-012-0212-2

Dhond, R. P., Yeh, C., Park, K., & Kettner, N. (2008). Acupuncture Modulates Resting State Connectivity in Default and Sensorimotor Brain Networks. *Pain*, *136*(3), 407–418.

Di, X., Zhu, S., Jin, H., Wang, P., Ye, Z., Zhou, K., … Rao, H. (2012). Altered resting brain function and structure in professional badminton players. *Brain Connectivity*, *2*(4), 225–33.

http://doi.org/10.1089/brain.2011.0050

Diamond, A., & Ling, D. S. (2016a). Conclusions about interventions, programs, and approaches for improving executive functions that appear justified and those that, despite much hype, do not. *Developmental Cognitive Neuroscience*, *18*, 34–48. http://doi.org/10.1016/j.dcn.2015.11.005

Diamond, A., & Ling, D. S. (2016b). Conclusions about interventions, programs, and approaches for improving executive functions that appear justified and those that, despite much hype, do not. *Developmental Cognitive Neuroscience*, *18*, 34–48. http://doi.org/10.1016/j.dcn.2015.11.005

Dosenbach, N. U. F., Fair, D. a, Miezin, F. M., Cohen, A. L., Wenger, K. K., Dosenbach, R. a T., … Petersen, S. E. (2007). Distinct brain networks for adaptive and stable task control in humans. *Proceedings of the National Academy of Sciences*, *104*(26), 11073–11078. http://doi.org/10.1073/pnas.0704320104

Doughty, M. J. (2001). Consideration of three types of spontaneous eyeblink activity in normal humans: during reading and video display terminal use, in primary gaze, and while in conversation. *Optometry and Vision Science : Official Publication of the American Academy of Optometry*, *78*(10), 712–725. http://doi.org/10.1097/00006324-200110000-00011

Draganski, B., & May, A. (2008). Training-induced structural changes in the adult human brain. *Behavioural Brain Research*, *192*(1), 137–142. http://doi.org/10.1016/j.bbr.2008.02.015

Dreisbach, G., Müller, J., Goschke, T., Strobel, A., Schulze, K., Lesch, K.-P., & Brocke, B. (2005). Dopamine and cognitive control: the influence of spontaneous eyeblink rate and dopamine gene polymorphisms on perseveration and distractibility. *Behavioral Neuroscience*, *119*(2), 483–490. http://doi.org/10.1037/0735-7044.119.2.483

Dumas, D., & Schmidt, L. (2015). Relational reasoning as predictor for engineering ideation success using TRIZ. *Journal of Engineering Design*, *26*(1–3), 74–88. http://doi.org/10.1080/09544828.2015.1020287

Dumontheil, I., Houlton, R., Christoff, K., & Blakemore, S.-J. (2010). Development of relational reasoning during adolescence. *Developmental Science*, *13*(6), F15–F24.

Duncan, J., Chylinski, D., Mitchell, D. J., & Bhandari, A. (2017). Complexity and compositionality in fluid intelligence. *Proceedings of the National Academy of Sciences*, *114*(20), 5295–5299. http://doi.org/10.1073/pnas.1621147114

Duncan, N. W., & Northoff, G. (2013). Overview of potential procedural and participant-related confounds for neuroimaging of the resting state. *Journal of Psychiatry & Neuroscience : JPN*, *38*(2), 84–96. http://doi.org/10.1503/jpn.120059

Eckstein, M. K., Guerra-carrillo, B., Singley, A. T. M., & Bunge, S. A. (2016). Beyond eye gaze: What else can eyetracking reveal about cognition and cognitive development? *Developmental Cognitive Neuroscience*, 1–23. http://doi.org/10.1016/j.dcn.2016.11.001

Einhäuser, W., Koch, C., & Carter, O. (2010). Pupil dilation betrays the timing of decisions. *Frontiers in Human Neuroscience*, *4*(18). http://doi.org/10.3389/fnhum.2010.00018

Eldar, E., Cohen, J. D., & Niv, Y. (2013). The effects of neural gain on attention and learning. *Nature Neuroscience*, *16*(8), 1146–53. http://doi.org/10.1038/nn.3428

Elsworth, J. D., Lawrence, M. S., Roth, R. H., Taylor, J. R., Mailman, R. B., Nichols, D. E., … Redmond, D. E. (1991). D1 and D2 dopamine receptors independently regulate spontaneous blink rate in the vervet monkey. *The Journal of Pharmacology and Experimental Therapeutics*, *259*, 595–600.

Ernst, M., Romeo, R. D., & Andersen, S. L. (2009). Neurobiology of the development of motivated behaviors in adolescence: A window into a neural systems model. *Pharmacology Biochemistry and Behavior*, *93*(3), 199–211. http://doi.org/10.1016/j.pbb.2008.12.013

Falch, T., & Sandgren Massih, S. (2011). The effect of education on cognitive ability. *Economic Inquiry*, *49*(3), 838–856. http://doi.org/10.1111/j.1465-7295.2010.00312.x

Feng, G. (2011). Eye Tracking: A Brief Guide for Developmental Researchers. *Journal of Cognition and Development*, *12*(1), 1–11. http://doi.org/10.1080/15248372.2011.547447

Ferreira, F., Apel, J., & Henderson, J. M. (2008). Taking a new look at looking at nothing. *Trends in Cognitive Sciences*, *12*(11), 405–410. http://doi.org/10.1016/j.tics.2008.07.007

Ferrer, E., & McArdle, J. J. (2004). An Experimental Analysis of Dynamic Hypotheses About Cognitive Abilities and Achievement From Childhood to Early Adulthood. *Developmental Psychology*, *40*(6), 935–952. http://doi.org/10.1037/0012-1649.40.6.935

Foote, S. L., Bloom, F. E., & Aston-Jones, G. (1983). Nucleus locus ceruleus: new evidence of anatomical and physiological specificity. *Physiological Reviews*, *63*(3), 844–914.

Foote, S. L., Freedman, R., & Oliver, A. P. (1975). Effects of putative neurotransmitters on neuronal activity in monkey auditory cortex. *Brain Research*, *86*(2), 229–242.

Franchak, J. M., Kretch, K. S., Soska, K. C., & Adolph, K. E. (2011). Head-Mounted Eye Tracking: A New Method to Describe Infant Looking: Head-Mounted Eye Tracking. *Child Development*, *82*(6), 1738–1750. http://doi.org/10.1111/j.1467-8624.2011.01670.x

Frank, M. J., Loughry, B., & O'Reilly, R. C. (2001). Interactions between frontal cortex and basal ganglia in working memory: A computational model. *Cognitive, Affective, & Behavioral Neuroscience*, *1*(2), 137–160.

French, R. M., & Thibaut, J.-P. (2014). Using eye-tracking to predict children's success or failure on analogy tasks. In *Livre/Conférence Proceedings of the Thirty-Sixth Annual Meeting of the Cognitive Science Society. Austin, TX: Cognitive Science Society* (pp. 2222–2227).

Fried, M., Tsitsiashvili, E., Bonneh, Y. S., Sterkin, A., Wygnanski-Jaffe, T., Epstein, T., & Polat, U. (2014). ADHD subjects fail to suppress eye blinks and microsaccades while anticipating visual stimuli but recover with medication. *Vision Research*, *101*, 62–72.

http://doi.org/10.1016/j.visres.2014.05.004

Fukuda, K. (2001). Eye blinks: New indices for the detection of deception. *International Journal of Psychophysiology*, *40*(June 2000), 239–245. http://doi.org/10.1016/S0167-8760(00)00192-6

Funahashi, S., Bruce, C. J., & Goldman-Rakic, P. S. (1989). Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *Journal of Neurophysiology*, *61*(2), 331–349.

Fuster, J. (2015). *The prefrontal cortex.* ACADEMIC PRESS INC.

Gabard-durnam, L. J., Gee, D. G., Goff, B., Flannery, J., Telzer, E., Humphreys, K. L., … Caldera, C. (2016). Stimulus-Elicited Connectivity Influences Resting-State Connectivity Years Later in Human Development : A Prospective Study. *The Journal of Neuroscience*, *36*(17), 4771–4784. http://doi.org/10.1523/JNEUROSCI.0598-16.2016

Gabrieli, J. D. E., Ghosh, S. S., & Whitfield-Gabrieli, S. (2015). Prediction as a humanitarian and pragmatic contribution from human cognitive neuroscience. *Neuron*, *85*(1), 11–26. http://doi.org/10.1016/j.neuron.2014.10.047

Galea, J. M., Vazquez, A., Pasricha, N., de Xivry, J.-J. O., & Celnik, P. (2011). Dissociating the roles of the cerebellum and motor cortex during adaptive learning: the motor cortex retains what the cerebellum learns. *Cerebral Cortex (New York, N.Y. : 1991)*, *21*(8), 1761–70. http://doi.org/10.1093/cercor/bhq246

Galván, A. (2010). Neural plasticity of development and learning. *Human Brain Mapping*, *31*(6), 879–890. http://doi.org/10.1002/hbm.21029

Gamino, J. F., Motes, M. M., Riddle, R., Lyon, G. R., Spence, J. S., & Chapman, S. B. (2014). Enhancing inferential abilities in adolescence: new hope for students in poverty. *Frontiers in Human Neuroscience*, *8*(December), 924. http://doi.org/10.3389/fnhum.2014.00924

Ganguli, M., Snitz, B. E., Lee, C.-W., Vanderbilt, J., Saxton, J., & Chang, C.-C. H. (2010). Age and education effects and norms on cognitive test battery from a population-based cohort: The Monongahela-Youghiogheny Healthy Aging Team (MYHAT). *Aging Ment Health*, *14*(1), 100–107. http://doi.org/10.1080/13607860903071014.Age

Gehricke, J.-G., Ornitz, E. M., & Siddarth, P. (2002). Differentiating between reflex and spontaneous blinks using simultaneous recording of the orbicularis oculi electromyogram and the electro-oculogram in startle research. *International Journal of Psychophysiology*, *44*(3), 261–268.

Gillebert, C. R., & Mantini, D. (2013). Functional Connectivity in the Normal and Injured Brain. *The Neuroscientist*, *19*(5), 509–522. http://doi.org/10.1177/1073858412463168

Gilzenrat, M. S., Nieuwenhuis, S., Jepma, M., & Cohen, J. D. (2010). Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function. *Cognitive, Affective & Behavioral Neuroscience*, *10*(2), 252–69. http://doi.org/10.3758/CABN.10.2.252

Glady, Y., Thibaut, J.-P., French, R. M., & Blaye, A. (2012). Explaining children's failure in analogy making tasks: A problem of focus of attention? In *Proceedings of the Annual Meeting of the Cognitive Science Meeting*.

Glimcher, P. W. (2011). Understanding dopamine and reinforcement learning: The dopamine reward prediction error hypothesis. *Proceedings of the National Academy of Sciences*, *108*(Supplement 3), 15647–15654.

Goel, V., & Dolan, R. J. (2004). Differential involvement of left prefrontal cortexin inductive and deductive reasoning. *Cognition*, *93*(3). http://doi.org/10.1016/j.cognition.2004.03.001

Goldin, A. P., Hermida, M. J., Shalom, D. E., Elias Costa, M., Lopez-Rosenfeld, M., Segretin, M. S., … Sigman, M. (2014). Far transfer to language and math of a short software-based gaming intervention. *Proceedings of the National Academy of Sciences*, *111*(17), 6443–6448. http://doi.org/10.1073/pnas.1320217111

Goldinger, S. D., & Papesh, M. H. (2012). Pupil Dilation Reflects the Creation and Retrieval of Memories. *Current Directions in Psychological Science*, *21*(2), 90–95. http://doi.org/10.1177/0963721412436811

Goldman-Rakic, P. S., Muly, E. C., & Williams, G. V. (2000). D(1) receptors in prefrontal cells and circuits. *Brain Research. Brain Research Reviews*, *31*(2–3), 295–301.

Goldstein, R. Z., & Volkow, N. D. (2011). Dysfunction of the prefrontal cortex in addiction: neuroimaging findings and clinical implications. *Nature Reviews. Neuroscience*, *12*(11), 652–69. http://doi.org/10.1038/nrn3119

Goodwin, G. P., & Johnson-Laird, P. N. (2006). Reasoning about the relations between relations. *Quarterly Journal of Experimental Psychology*, *59*(6), 1047–1069. http://doi.org/10.1080/02724980543000169

Granholm, E., Asarnow, R. F., Sarkin, A. J., & Dykes, K. L. (1996). Pupillary responses index cognitive resource limitations. *Psychophysiology*, *33*(4), 457–461. http://doi.org/10.1111/j.1469-8986.1996.tb01071.x

Granholm, E., & Steinhauer, S. R. (2004). Pupillometric measures of cognitive and emotional processes. *International Journal of Psychophysiology*, *52*(1), 1–6. http://doi.org/10.1016/j.ijpsycho.2003.12.001

Grant, E. R., & Spivey, M. J. (2003). Eye movements and problem solving guiding attention guides thought. *Psychological Science*, *14*(5), 462–466.

Gredebäck, G., Johnson, S., & von Hofsten, C. (2009). Eye Tracking in Infancy Research. *Developmental Neuropsychology*, *35*(1), 1–19. http://doi.org/10.1080/87565640903325758

Green, H. J., Lemaire, P., & Dufau, S. (2007). Eye movement correlates of younger and older adults' strategies for complex addition. *Acta Psychologica*, *125*(3), 257–278. http://doi.org/10.1016/j.actpsy.2006.08.001

Greene, R. L. (1989). Spacing effects in memory: Evidence for a two-process account. *Journal* *of Experimental Psychology: Learning, Memory, and Cognition*, *15*(3), 371–377. http://doi.org/10.1037/0278-7393.15.3.371

Greenough, W. T., Black, J. E., Wallace, C. S., Development, C., & Jun, N. (1987). Experience and Brain Development. *Child Development*, *58*(3), 539–559. http://doi.org/10.2307/1130197

Groman, S. M., James, a S., Seu, E., Tran, S., Clark, T. a, Harpster, S. N., … Jentsch, J. D. (2014). In the blink of an eye: relating positive-feedback sensitivity to striatal dopamine D2-like receptors through blink rate. *J Neurosci*, *34*(43), 14443–14454. http://doi.org/10.1523/JNEUROSCI.3037-14.2014

Grosbras, M.-H., Laird, A. R., & Paus, T. (2005). Cortical regions involved in eye movements, shifts of attention, and gaze perception. *Human Brain Mapping*, *25*(1), 140–154. http://doi.org/10.1002/hbm.20145

Guipponi, O., Odouard, S., Pinède, S., Wardak, C., & Ben Hamed, S. (2014). fMRI Cortical Correlates of Spontaneous Eye Blinks in the Nonhuman Primate. *Cerebral Cortex (New York, N.Y. : 1991)*, 1–2. http://doi.org/10.1093/cercor/bhu038

Gujar, N., Yoo, S., Hu, P., & Walker, M. P. (2009). The Unrested Resting Brain: Sleep Deprivation Alters Activity within the Default-mode Network. *Journal of Cognitive Neuroscience*, 1–12.

Gütl, C., Pivec, M., Trummer, C., García-Barrios, V. M., Mödritscher, F., Pripfl, J., & Umgeher, M. (2005). Adele (adaptive e-learning with eye-tracking): Theoretical background, system architecture and application scenarios. *European Journal of Open, Distance and E-Learning*, *8*(2).

Halford, G. S., Wilson, W. H., & Phillips, S. (1998). Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. *Behavioral and Brain Sciences*, *21*(6), 803–864. http://doi.org/10.1017/S0140525X98001769

Halford, G. S., Wilson, W. H., & Phillips, S. (2010). Relational knowledge: the foundation of higher cognition. *Trends in Cognitive Sciences*, *14*(11), 497–505. http://doi.org/10.1016/j.tics.2010.08.005

Hallett, M. (2000). Clinical physiology of dopa dyskinesia. *Annals of Neurology*, *47*(4 Suppl 1), S147-150; discussion S150-153.

Hallett, P. E. (1978). Primary and secondary saccades to goals defined by instructions. *Vision Research*, *18*(10), 1279–1296.

Hampshire, A., Highfield, R. R., Parkin, B. L., & Owen, A. M. (2012). Fractionating Human Intelligence. *Neuron*, *76*(6), 1225–1237. http://doi.org/10.1016/j.neuron.2012.06.022

Han, K., Davis, R. A., Chapman, S. B., & Krawczyk, D. C. (2017). Strategy-based reasoning training modulates cortical thickness and resting-state functional connectivity in adults with chronic traumatic brain injury. *Brain and Behavior*, *7*(5), 1–27.

http://doi.org/10.1002/brb3.687

Hannula, D. E., Althoff, R. R., Warren, D. E., Riggs, L., Cohen, N. J., & Ryan, J. D. (2010). Worth a glance: using eye movements to investigate the cognitive neuroscience of memory. *Frontiers in Human Neuroscience*, *4*(October), 166. http://doi.org/10.3389/fnhum.2010.00166

Hansen, K. T., Heckman, J. J., & Mullen, K. (2003). The Effect of Schooling and Ability on Achievement Test Scores. *IZA Discussion Paper Series*, (826).

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. http://doi.org/10.1371/journal.pone.0134467

Härnqvist, K. (1968). Relative Changes in Intelligence from 13 to 18. *Scandinavian Journal of Psychology*, *9*, 50–82.

Hartshorne, J. K., & Germine, L. T. (2015). When does cognitive functioning peak? The asynchronous rise and fall of different cognitive abilities across the life span. *Psychological Science*, *26*(4), 433–43. http://doi.org/10.1177/0956797614567339

Harvey, L. (2000). New realities: The relationship between higher education and employment. *Tertiary Education and Management*, *6*(1), 3–17. http://doi.org/10.1080/13583883.2000.9967007

Hayes, T. R., Petrov, A. A., & Sederberg, P. B. (2011). A novel method for analyzing sequential eye movements reveals strategic influence on Raven's Advanced Progressive Matrices. *Journal of Vision*, *11*(10), 10–10. http://doi.org/10.1167/11.10.10

Hebb, D. (1949). *The Organisation of Behaviour: A Neuropsychological Theory*. New York: John Wiley & Sons.

Hensch, T. K. (2005). Critical period plasticity in local cortical circuits. *Nature Reviews. Neuroscience*, *6*(11), 877–88. http://doi.org/10.1038/nrn1787

Hepach, R., & Westermann, G. (2016). Pupillometry in Infancy Research. *Journal of Cognition and Development*, *17*(3), 359–377. http://doi.org/10.1080/15248372.2015.1135801

Herlenius, E., & Lagercrantz, H. (2001). Neurotransmitters and neuromodulators during early human development. *Early Human Development*, *65*(1), 21–37.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures.* Chicago: OUP Oxford.

Hong, L., Walz, J. M., & Sajda, P. (2014). Your Eyes Give You Away: Prestimulus Changes in Pupil Diameter Correlate with Poststimulus Task-Related EEG Dynamics. *PLOS ONE*, *9*(3), e91321. http://doi.org/10.1371/journal.pone.0091321

Hossain, G., & Yeasin, M. (2014). Understanding Effects of Cognitive Load from Pupillary Responses Using Hilbert Analytic Phase. In *2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops* (pp. 381–386). http://doi.org/10.1109/CVPRW.2014.62

Huestegge, L., Radach, R., Corbic, D., & Huestegge, S. M. (2009). Oculomotor and linguistic determinants of reading development: A longitudinal study. *Vision Research*, *49*(24), 2948–2959. http://doi.org/10.1016/j.visres.2009.09.012

Jackson, I., & Sirois, S. (2009). Infant cognition: Going full factorial with pupil dilation. *Developmental Science*, *12*(4), 670–679. http://doi.org/10.1111/j.1467-7687.2008.00805.x

Jiang, X., Tien, G., Huang, D., Zheng, B., & Atkins, M. S. (2013). Capturing and evaluating blinks from video-based eyetrackers. *Behavior Research Methods*, *45*(3), 656–663. http://doi.org/10.3758/s13428-012-0294-x

Johansson, R., Holsanova, J., & Holmqvist, K. (2006). Pictures and spoken descriptions elicit similar eye movements during mental imagery, both in light and in complete darkness. *Cognitive Science*, *30*(6), 1053–1079.

Johnson-Laird, P. N. (2010). Mental models and human reasoning. *Proceedings of the National Academy of Sciences*, *107*(43), 18243–18250. http://doi.org/10.1073/pnas.1012933107

Johnson, C., & Wilbrecht, L. (2011). Juvenile mice show greater flexibility in multiple choice reversal learning than adults. *Developmental Cognitive Neuroscience*, *1*(4), 540–551. http://doi.org/10.1016/j.dcn.2011.05.008

Johnson, E. L., Miller Singley, A. T., Peckham, A. D., Johnson, S. L., & Bunge, S. A. (2014). Task-evoked pupillometry provides a window into the development of short-term memory capacity. *Frontiers in Psychology*, *5*(MAR), 1–8. http://doi.org/10.3389/fpsyg.2014.00218

Johnson, S. P., Amso, D., & Slemmer, J. A. (2003). Development of object concepts in infancy: Evidence for early learning in an eye-tracking paradigm. *Proceedings of the National Academy of Sciences*, *100*(18), 10568–10573.

Jongkees, B. J., & Colzato, L. S. (2016). Spontaneous eye blink rate as predictor of dopamine-related cognitive function—A review. *Neuroscience & Biobehavioral Reviews*, *71*, 58–82. http://doi.org/10.1016/j.neubiorev.2016.08.020

Joshi, S., Li, Y., Kalwani, R., & Gold, J. I. (2017). Relationships between pupil diameter and neuronal activity in the locus coeruleus, colliculi, and cingulate cortex. *Neuron*, *89*(1), 221–234. http://doi.org/10.1016/j.neuron.2015.11.028.Relationships

Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive Psychology*, *8*(4), 441–480.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, *87*(4), 329.

Just, M. A., Carpenter, P. A., & Miyake, A. (2003). Neuroindices of cognitive workload:

Neuroimaging, pupillometric and event-related potential studies of brain work. *Theoretical Issues in Ergonomics Science*, *4*(1–2), 56–88. http://doi.org/10.1080/14639220210159735

Kahneman, D., & Beatty, J. (1966). Pupil Diameter and Load on Memory. *Science*, *154*(3756), 1583–1585. http://doi.org/10.1126/science.154.3756.1583

Kail, R. V, Lerv, A., & Hulme, C. (2016). Longitudinal evidence linking processing speed to the development of reasoning. *Developmental Science*, *19*(6), 1067–1074. http://doi.org/10.1111/desc.12352

Kaminer, J., Powers,  a. S., Horn, K. G., Hui, C., & Evinger, C. (2011). Characterizing the Spontaneous Blink Generator: An Animal Model. *Journal of Neuroscience*, *31*(31), 11256–11267. http://doi.org/10.1523/JNEUROSCI.6218-10.2011

Kamp, S.-M., & Donchin, E. (2015). ERP and pupil responses to deviance in an oddball paradigm. *Psychophysiology*, *52*(4), 460–471. http://doi.org/10.1111/psyp.12378

Karatekin, C., Marcus, D. J., & Couperus, J. W. (2007). Regulation of cognitive resources during sustained attention and working memory in 10-year-olds and adults. *Psychophysiology*, *44*(1), 128–44. http://doi.org/10.1111/j.1469-8986.2006.00477.x

Karson, C. (1983). Spontaneous eye-blink rates and dopaminergic systems. *Brain*, *106*(3), 643–653.

Karson, C. N. (1988). Physiology of normal and abnormal blinking. *Advances in Neurology*, *49*, 25–37.

Karson, C. N., Berman, K. F., Donnelly, E. F., Mendelson, W. B., Kleinman, J. E., & Wyatt, R. J. (1981). Speaking, thinking, and blinking. *Psychiatry Research*, *5*, 243–246. http://doi.org/10.1016/0165-1781(81)90070-6

Kaufman, A. S., Kaufman, J. C., Liu, X., & Johnson, C. K. (2009). How do educational attainment and gender relate to fluid intelligence, crystallized intelligence, and academic skills at ages 22-90 years? *Archives of Clinical Neuropsychology*, *24*(2), 153–163. http://doi.org/10.1093/arclin/acp015

Klauer, K. C., Meiser, T., & Naumer, B. (2000). Training propositional reasoning. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *53*(3), 868–895. http://doi.org/10.1080/713755911

Klauer, K. J., & Phye, G. D. (2008). Inductive Reasoning: A Training Approach. *Review of Educational Research*, *78*(1), 85–123. http://doi.org/10.3102/0034654307313402

Klauer, K. J., Willmes, K., & Phye, G. D. (2002). Inducing Inductive Reasoning: Does It Transfer to Fluid Intelligence? *Contemporary Educational Psychology*, *27*(1), 1–25. http://doi.org/10.1006/ceps.2001.1079

Kleven, M. S., & Koek, W. (1996). Differential effects of direct and indirect dopamine agonists on eye blink rate in cynomolgus monkeys. *The Journal of Pharmacology and Experimental Therapeutics*, *279*, 1211–1219. http://doi.org/10.1097/00008877-199605001-00128

Klingberg, T., Fernell, E., Olesen, P. J., Johnson, M., Gustafsson, P., Dahlström, K., … Westerberg, H. (2005). Computerized training of working memory in children with ADHD--a randomized, controlled trial. *Journal of the American Academy of Child and Adolescent Psychiatry*, *44*(2), 177–86. http://doi.org/10.1097/00004583-200502000-00010

Klingner, C. M., Hasler, C., Brodoehl, S., Axer, H., & Witte, O. W. (2012). Perceptual plasticity is mediated by connectivity changes of the medial thalamic nucleus. *Human Brain Mapping*, *000*. http://doi.org/10.1002/hbm.22074

Klingner, J. (2010). *Measuring cognitive load during visual tasks by combining pupillometry and eye tracking*.

Klingner, J., Tversky, B., & Hanrahan, P. (2011). Effects of visual and verbal presentation on cognitive load in vigilance, memory, and arithmetic tasks. *Psychophysiology*, *48*(3), 323–332. http://doi.org/10.1111/j.1469-8986.2010.01069.x

Knoll, L. J., Fuhrmann, D., Sakhardande, A. L., Stamp, F., Speekenbrink, M., & Blakemore, S.-J. (2016). A Window of Opportunity for Cognitive Training in Adolescence. *Psychological Science*, *27*(12), 1620–1631. http://doi.org/10.1177/0956797616671327

Kolb, B., Mychasiuk, R., Muhammad, A., Li, Y., Frost, D. O., & Gibb, R. (2012). Experience and the developing prefrontal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *109 Suppl*, 17186–93. http://doi.org/10.1073/pnas.1121251109

Kotani, M., Kiyoshi, A., Murai, T., Nakako, T., Matsumoto, K., Matsumoto, A., … Ikeda, K. (2016). The dopamine D1 receptor agonist SKF-82958 effectively increases eye blinking count in common marmosets. *Behavioural Brain Research*, *300*, 25–30. http://doi.org/10.1016/j.bbr.2015.11.028

Kowler, E. (2011). Eye movements: The past 25years. *Vision Research*, *51*(13), 1457–1483. http://doi.org/10.1016/j.visres.2010.12.014

Krawczyk, D. C. (2012). The cognition and neuroscience of relational reasoning. *Brain Research*, *1428*, 13–23. http://doi.org/10.1016/j.brainres.2010.11.080

Królak, A., & Strumiłło, P. (2012). Eye-blink detection system for human–computer interaction. *Universal Access in the Information Society*, *11*(4), 409–419. http://doi.org/10.1007/s10209-011-0256-6

Kühn, S., Gleich, T., Lorenz, R. C., Lindenberger, U., & Gallinat, J. (2014). Playing super mario induces structural brain plasticity: Gray matter changes resulting from training with a commercial video game. *Molecular Psychiatry*, *19*(2), 265–271. http://doi.org/10.1038/mp.2013.120

Kühn, S., & Lindenberger, U. (2016). Research on Human Plasticity in Adulthood: A Lifespan Agenda. In W. Schaie & S. Willis (Eds.), *Handbook of the Psychology of Aging* (8th ed., pp. 105–123). San Diego: Elsevier Inc. http://doi.org/10.1016/B978-0-12-411469-2.00006-6

Laeng, B., Ørbo, M., Holmlund, T., & Miozzo, M. (2010). Pupillary Stroop effects. *Cognitive*

*Processing*, *12*(1), 13–21. http://doi.org/10.1007/s10339-010-0370-z

Laeng, B., Sirois, S., & Gredeback, G. (2012). Pupillometry: A Window to the Preconscious? *Perspectives on Psychological Science*, *7*(1), 18–27. http://doi.org/10.1177/1745691611427305

Lai, M.-L., Tsai, M.-J., Yang, F.-Y., Hsu, C.-Y., Liu, T.-C., Lee, S. W.-Y., … Tsai, C.-C. (2013). A review of using eye-tracking technology in exploring learning from 2000 to 2012. *Educational Research Review*, *10*, 90–115. http://doi.org/10.1016/j.edurev.2013.10.001

Lavezzo, M. M., Schellini, S. A., Padovani, C. R., & Hirai, F. E. (2007). Eye blink in newborn and preschool-age children. *Acta Ophthalmologica*, *86*(3), 275–278. http://doi.org/10.1111/j.1600-0420.2007.00969.x

Levin, L. A., Nilsson, S. F. E., Ver Hoeve, J., Wu, S., Kaufman, P. L., & Alm, A. (Eds.). (2011). *Adler's Physiology of the Eye*. Elsevier Health Sciences.

Lewkowicz, D. J., & Hansen-Tift, A. M. (2012). Infants deploy selective attention to the mouth of a talking face when learning speech. *Proceedings of the National Academy of Sciences*, *109*(5), 1431–1436. http://doi.org/10.1073/pnas.1114783109

Lindenberger, U., Wenger, E., & Lövdén, M. (2017). Towards a stronger science of human plasticity. *Nature Reviews Neuroscience*, *18*(5), 261–262. http://doi.org/10.1038/nrn.2017.44

Lisman, J., Schulman, H., & Cline, H. (2002). The molecular basis of CaMKII function in synaptic and behavioural memory. *Nature Reviews. Neuroscience*, *3*(3), 175–90. http://doi.org/10.1038/nrn753

Liu, C., Li, W., Tong, K. A., & Yeom, K. W. (2015). Susceptibility-Weighted Imaging and Quantitative Susceptibility Mapping in the Brain. *J Magn Reson Imaging*, *42*(1), 23–41. http://doi.org/10.1002/jmri.24768.Susceptibility-Weighted

Loewenfeld, I. E., & Lowenstein, O. (1993). *The Pupil: Anatomy, Physiology, and Clinical Applications* (Vol. 2). Detroit, MI: Iowa State University Press.

Lövdén, M., Bäckman, L., Lindenberger, U., Schaefer, S., & Schmiedek, F. (2010a). A theoretical framework for the study of adult cognitive plasticity. *Psychological Bulletin*, *136*(4), 659–76. http://doi.org/10.1037/a0020080

Lövdén, M., Bäckman, L., Lindenberger, U., Schaefer, S., & Schmiedek, F. (2010b). A theoretical framework for the study of adult cognitive plasticity. *Psychological Bulletin*, *136*(4), 659–676. http://doi.org/10.1037/a0020080

Löwenstein, O. (1920). Experimentelle Beiträge zur Lehre von den katatonischen Pupillenveränderungen. *European Neurology*, *47*(4), 194–215. http://doi.org/10.1159/000190690

Lowenstein, O., & Loewenfeld, I. E. (1958). Electronic pupillography: A new instrument and some clinical applications. *A.M.A. Archives of Ophthalmology*, *59*(3), 352–363.

http://doi.org/10.1001/archopht.1958.00940040058007

Luna, B., Garver, K. E., Urban, T. A., Lazar, N. A., & Sweeney, J. A. (2004). Maturation of cognitive processes from late childhood to adulthood. *Child Development*, *75*(5), 1357–1372.

Luna, B., Marek, S., Larsen, B., Tervo-Clemmens, B., & Chahal, R. (2015). An Integrative Model of the Maturation of Cognitive Control. *Annual Review of Neuroscience*, *38*(1), 151–170. http://doi.org/10.1146/annurev-neuro-071714-034054

Luna, B., & Velanova, K. (2011). Development from reflexive to controlled eye movements. In S. P. Liversedge, I. Gilchrist, & S. Everling (Eds.), *The Oxford Handbook of Eye Movements* (1st ed.). Oxford University Press.

Luna, B., Velanova, K., & Geier, C. F. (2008). Development of eye-movement control. *Brain and Cognition*, *68*(3), 293–308. http://doi.org/10.1016/j.bandc.2008.08.019

Luo, C., Guo, Z., Lai, Y., Liao, W., Liu, Q., Kendrick, K. M., … Li, H. (2012). Musical Training Induces Functional Plasticity in Perceptual and Motor Networks: Insights from Resting-State fMRI. *PLoS ONE*, *7*(5), e36568. http://doi.org/10.1371/journal.pone.0036568

Lustig, C., Shah, P., Seidler, R., & Reuter-Lorenz, P. A. (2009). Aging, Training, and the Brain: A Review and Future Directions. *Neuropsychology Review*, *19*(4), 504–522. http://doi.org/10.1007/s11065-009-9119-9

Ma, L., Narayana, S., Robin, D. A., Fox, P. T., & Xiong, J. (2012). Changes Occur in Resting State Network of Motor System during 4 Weeks of Motor Skill Learning. *NeuroImage*, *58*(1), 226–233. http://doi.org/10.1016/j.neuroimage.2011.06.014.Changes

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), 582–90. http://doi.org/10.1111/j.1467-7687.2010.01005.x

Mackey, A. P., Miller Singley, A. T., & Bunge, S. a. (2013). Intensive reasoning training alters patterns of brain connectivity at rest. *The Journal of Neuroscience*, *33*(11), 4796–803. http://doi.org/10.1523/JNEUROSCI.4141-12.2013

Mackey, A. P., Miller Singley, A. T., Wendelken, C., & Bunge, S. A. (2015). Characterizing behavioral and brain changes associated with practicing reasoning skills. *PLoS ONE*, *10*(9), 1–18. http://doi.org/10.1371/journal.pone.0137627

Mackey, A. P., Whitaker, K. J., & Bunge, S. A. (2012a). Experience-dependent plasticity in white matter microstructure: reasoning training alters structural connectivity. *Frontiers in Neuroanatomy*, *6*(August), 1–9. http://doi.org/10.3389/fnana.2012.00032

Mackey, A. P., Whitaker, K. J., & Bunge, S. a. (2012b). Experience-dependent plasticity in white matter microstructure: reasoning training alters structural connectivity. *Frontiers in Neuroanatomy*, *6*(August), 32. http://doi.org/10.3389/fnana.2012.00032

MacLachlan, C., & Howland, H. C. (2002). Normal values and standard deviations for pupil diameter and interpupillary distance in subjects aged 1 month to 19 years. *Ophthalmic and Physiological Optics*, *22*(3), 175–182. http://doi.org/10.1046/j.1475-1313.2002.00023.x

Marshall, K. C., Christie, M. J., Finlayson, P. G., & Williams, J. T. (1991). Developmental aspects of the locus coeruleus-noradrenaline system. *Progress in Brain Research*, *88*, 173–185.

Martin, J., Schönbacher, T., Nischelwitzer, A. K., Sprung, G., Pauschenwein, J., Sandtner, H., … Behmel, A. (2007). Tobii Eye Tracker Extension for Presentation. Graz: Vision Space. Retrieved from http://idk.fh-joanneum.at/2015/09/22/visionspace/

May, A. (2011). Experience-dependent structural plasticity in the adult human brain. *Trends in Cognitive Sciences*, *15*(10), 475–482. http://doi.org/10.1016/j.tics.2011.08.002

Mcardle, J. J., Ferrer-caja, E., Hamagami, F., & Woodcock, R. W. (2002). Comparative Longitudinal Structural Analyses of the Growth and Decline of Multiple Intellectual Abilities Over the Life Span. *Developmental Psychology*, *38*(1), 115–142. http://doi.org/10.1037//0012-1649.38.1.115

McArdle, J. J., Ferrer-Caja, E., Hamagami, F., & Woodcock, R. W. (2002). Comparative longitudinal structural analyses of the growth and decline of multiple intellectual abilities over the life span. *Developmental Psychology*, *38*(1), 115–142. http://doi.org/10.1037/0012-1649.38.1.115

McEwen, B. S., & Morrison, J. H. (2013). The Brain on Stress: Vulnerability and Plasticity of the Prefrontal Cortex over the Life Course. *Neuron*, *79*(1), 16–29. http://doi.org/10.1016/j.neuron.2013.06.028

Mcnab, F., Nix, P., Hauth, L., Jorgensen, E. M., & Bastiani, M. (2009). Changes in Cortical Dopamine D1 Receptor Binding Associated with Cognitive Training. *Science*, *323*(800), 1095–1099. http://doi.org/10.1126/science.1229223

Meindertsma, T. (2014). *Pupil dilation as a marker for neuromodulation: How strong is the evidence for the claim that pupil dilation reflects central noradrenaline release from the locus coeruleus?* Leiden University.

Melby-Lervåg, M., & Hulme, C. (2013b). Is working memory training effective? A meta-analytic review. *Developmental Psychology*, *49*(2), 270–291. http://doi.org/10.1037/a0028228

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

Miall, R. C., & Robertson, E. M. (2006). Functional Imaging: Is the Resting Brain Resting? *Current Biology*, *16*(23), R998–R1000. http://doi.org/10.1016/j.cub.2006.10.041

Money, K. M., & Stanwood, G. D. (2013). Developmental origins of brain disorders: roles for dopamine. *Frontiers in Cellular Neuroscience*, *7*. http://doi.org/10.3389/fncel.2013.00260

Moreau, D., & Conway, A. R. a. a. (2014). The case for an ecological approach to cognitive training. *Trends in Cognitive Sciences*, *18*(7), 334–336. http://doi.org/10.1016/j.tics.2014.03.009

Morey, R. D., Rouder, J. N., & Jamil, T. (2015). Computation of Bayes Factors for Common Designs, 54. Retrieved from http://bayesfactorpcl.r-forge.r-project.org/%0ABugReports

Morrison, G. E., Simone, C. M., Ng, N. F., & Hardy, J. L. (2015). Reliability and validity of the NeuroCognitive Performance Test , a web-based neuropsychological assessment. *Frontiers in Psychology*, *6*(November), 1–15. http://doi.org/10.3389/fpsyg.2015.01652

Mottershead, J. P., Schmierer, K., Clemence, M., Thornton, J. S., Scaravilli, F., Barker, G. J., … Miller, D. H. (2003). High field MRI correlates of myelin content and axonal density in multiple sclerosis: A post-mortem study of the spinal cord. *Journal of Neurology*, *250*(11), 1293–1301. http://doi.org/10.1007/s00415-003-0192-3

Müller, J., Dreisbach, G., Brocke, B., Lesch, K. P., Strobel, A., & Goschke, T. (2007). Dopamine and cognitive control: The influence of spontaneous eyeblink rate, DRD4 exon III polymorphism and gender on flexibility in set-shifting. *Brain Research*, *1131*, 155–162. http://doi.org/10.1016/j.brainres.2006.11.002

Müller, J., Philiastides, M. G., & Newsome, W. T. (2005). Microstimulation of the superior colliculus focuses attention without moving the eyes. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(3), 524–529. http://doi.org/10.1073/pnas.0408311101

Munoz, D. P., & Everling, S. (2004). Look away: the anti-saccade task and the voluntary control of eye movement. *Nature Reviews Neuroscience*, *5*(3), 218–228. http://doi.org/10.1038/nrn1345

Murphy, K., Birn, R. M., Handwerker, D. A., Jones, T. B., & Bandettini, P. A. (2009). The impact of global signal regression on resting state correlations: Are anti-correlated networks introduced? *NeuroImage*, *44*(3), 893–905. http://doi.org/10.1016/j.neuroimage.2008.09.036

Murphy, P. R., O'Connell, R. G., O'Sullivan, M., Robertson, I. H., & Balsters, J. H. (2014). Pupil diameter covaries with BOLD activity in human locus coeruleus. *Human Brain Mapping*, *35*(8), 4140–4154. http://doi.org/10.1002/hbm.22466

Murphy, P. R., Robertson, I. H., Balsters, J. H., & O'Connell, R. G. (2011). Pupillometry and P3 index the locus coeruleus-noradrenergic arousal function in humans: Indirect markers of locus coeruleus activity. *Psychophysiology*, *48*(11), 1532–1543. http://doi.org/10.1111/j.1469-8986.2011.01226.x

Neville, H. J., Stevens, C., Pakulak, E., Bell, T. a, Fanning, J., Klein, S., & Isbell, E. (2013). Family-based training program improves brain function, cognition, and behavior in lower socioeconomic status preschoolers. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(29), 12138–43. http://doi.org/10.1073/pnas.1304437110

Nieuwenhuis, S., Aston-Jones, G., & Cohen, J. D. (2005). Decision making, the P3, and the locus coeruleus--norepinephrine system. *Psychological Bulletin*, *131*(4), 510–532.

Noack, H., Lövdén, M., & Schmiedek, F. (2014). On the validity and generality of transfer effects in cognitive training research. *Psychological Research*, *78*(6), 773–789. http://doi.org/10.1007/s00426-014-0564-6

Noble, K. G., Korgaonkar, M. S., Grieve, S. M., & Brickman, A. M. (2013). Higher education is an age-independent predictor of white matter integrity and cognitive control in late adolescence. *Developmental Science*, *16*(5), 653–64. http://doi.org/10.1111/desc.12077

Nowak, W., Andrzej, H., & Kasprzak, H. (2007). Time-frequency analysis of spontaneous fluctuation of the pupil size of the human eye. *Optica Applicata*, *2*(38).

OECD. (2016). *Education at a Glance 2016: OECD Indicators*. Paris: OECD Publishing. http://doi.org/10.1787/eag-2016-en

Oh, J., Han, M., Peterson, B. S., & Jeong, J. (2012). Spontaneous eyeblinks are correlated with responses during the stroop task. *PLoS ONE*, *7*(4). http://doi.org/10.1371/journal.pone.0034871

Owen, A. M., Hampshire, A., Grahn, J. A., Stenton, R., Dajani, S., Burns, A. S., … Ballard, C. G. (2010a). Putting brain training to the test. *Nature*, *465*(7299), 775–778. http://doi.org/10.1038/nature09042

Owen, A. M., Hampshire, A., Grahn, J. A., Stenton, R., Dajani, S., Burns, A. S., … Ballard, C. G. (2010b). Putting brain training to the test. *Nature*, *465*(7299), 775–778. http://doi.org/10.1038/nature09042

Padmanabhan, A., & Luna, B. (2014). Developmental imaging genetics: Linking dopamine function to adolescent behavior. *Brain and Cognition*, *89*, 27–38. http://doi.org/10.1016/j.bandc.2013.09.011

Pas, P., Custers, R., Bijleveld, E., & Vink, M. (2014). Effort responses to suboptimal reward cues are related to striatal dopaminergic functioning. *Motivation and Emotion*, *38*(6), 759–770. http://doi.org/10.1007/s11031-014-9434-1

Patriat, R., Molloy, E. K., Meier, T. B., Kirk, G. R., Nair, V. A., Meyerand, M. E., … Birn, R. M. (2013). The effect of resting condition on resting-state fMRI reliability and consistency: A comparison between resting with eyes open, closed, and fixated. *NeuroImage*, *78*, 463–473. http://doi.org/10.1016/j.neuroimage.2013.04.013

Peckham, A. D., & Johnson, S. L. (2015). Spontaneous Eye-Blink Rate as an Index of Reward Responsivity Validation and Links to Bipolar Disorder. *Clinical Psychological Science*, *4*(3), 451–463.

Pedrotti, M., Lei, S., Dzaack, J., & Rötting, M. (2011). A data-driven algorithm for offline pupil signal preprocessing and eyeblink detection in low-speed eye-tracking protocols. *Behavior Research Methods*, *43*(2), 372–383. http://doi.org/10.3758/s13428-010-0055-7

Pessiglione, M., Seymour, B., Flandin, G., Dolan, R. J., & Frith, C. D. (2006). Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature*, *442*(7106), 1042–1045. http://doi.org/10.1038/nature05051

Poldrack, R. A. (2015). Is “efficiency” a useful concept in cognitive neuroscience? *Developmental Cognitive Neuroscience*, *11*, 12–17. http://doi.org/10.1016/j.dcn.2014.06.001

Pomplun, M., Sunkara, S., Fairley, A. V., & Xiao, M. (2009). Using Pupil Size as a Measure of Cognitive Workload in Video-Based Eye-Tracking Studies.

Ponder, E., & Kennedy, W. P. (1927). On the act of blinking. *Quarterly Journal of Experimental Physiology*, *18*(2), 89–110.

Powers, A. R., Hevey, M. a, & Wallace, M. T. (2012). Neural Correlates of Multisensory Perceptual Learning. *Journal of Neuroscience*, *32*(18), 6263–6274. http://doi.org/10.1523/JNEUROSCI.6138-11.2012

Prado, J., Chadha, A., & Booth, J. R. (2011). The brain network for deductive reasoning: a quantitative meta-analysis of 28 neuroimaging studies. *Journal of Cognitive Neuroscience*, *23*(11), 3483–97. http://doi.org/10.1162/jocn_a_00063

Puig, M. V., Rose, J., Schmidt, R., & Freund, N. (2014). Dopamine modulation of learning and memory in the prefrontal cortex: insights from studies in primates, rodents, and birds. *Frontiers in Neural Circuits*, *8*(August), 93. http://doi.org/10.3389/fncir.2014.00093

Qi, X. L., Meyer, T., Stanford, T. R., & Constantinidis, C. (2011). Changes in prefrontal neuronal activity after learning to perform a spatial working memory task. *Cerebral Cortex*, *21*(12), 2722–2732. http://doi.org/10.1093/cercor/bhr058

Raichle, M. E., MacLeod, a M., Snyder, a Z., Powers, W. J., Gusnard, D. a, & Shulman, G. L. (2001). A default mode of brain function. *Proceedings of the National Academy of Sciences*, *98*(2), 676–682. http://doi.org/10.1073/pnas.98.2.676

Rajkowski, J., Kubiak, P., & Aston-Jones, G. (1993). Correlations between locus coeruleus (LC) neural activity, pupil diameter and behavior in monkey support a role of LC in attention. In *Society for neuroscience abstracts* (Vol. 19, p. 974).

Ramos, B. P., & Arnsten, A. F. T. (2007). Adrenergic pharmacology and cognition: Focus on the prefrontal cortex. *Pharmacology & Therapeutics*, *113*(3), 523–536. http://doi.org/10.1016/j.pharmthera.2006.11.006

Ramsay, J. (2016, February). Functional Data Analysis.

Ramsay, J., & Silverman, B. W. (2002). *Applied functional data analysis: methods and case studies*. New York: Springer.

Rao, R. P. N., Zelinsky, G. J., Hayhoe, M. M., & Ballard, D. H. (2002). Eye movements in iconic visual search. *Vision Research*, *42*(11), 1447–1463. http://doi.org/10.1016/S0042-6989(02)00040-8

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372–422.

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology*, *62*(8), 1457–1506. http://doi.org/10.1080/17470210902816461

Rehder, B., & Hoffman, A. B. (2005). Eyetracking and selective attention in category learning. *Cognitive Psychology*, *51*(1), 1–41. http://doi.org/10.1016/j.cogpsych.2004.11.001

Richards, J. E., & Hunter, S. K. (2002). Testing neural models of the development of infant visual attention. *Developmental Psychobiology*, *40*(3), 226–236.

Richmond, J., & Nelson, C. A. (2009). Relational memory during infancy: evidence from eye tracking. *Developmental Science*, *12*(4), 549–556. http://doi.org/10.1111/j.1467-7687.2009.00795.x

Ritchie, S. J., Bates, T. C., & Deary, I. J. (2015). Is Education Associated With Improvements in General Cognitive Ability, or in Specific Skills ? *Developmental Psychology*, *51*(5), 573–582. http://doi.org/http://dx.doi.org/10.1037/a0038981

Ritchie, S. J., Bates, T. C., Der, G., Starr, J. M., & Deary, I. J. (2013). Education is associated with higher later life IQ scores, but not with faster cognitive processing speed. *Psychology and Aging*, *28*(2), 515–21. http://doi.org/10.1037/a0030820

Ritchie, S., & Tucker-Drob, E. (2017). How much does education improve intelligence? A meta-analysis. *PsyarXiv*, (November). http://doi.org/10.17605/osf.io/kymhp

Rjosk, V., Lepsien, J., Kaminski, E., Hoff, M., Sehm, B., Steele, C. J., … Ragert, P. (2017). Neural Correlates of Mirror Visual Feedback-Induced Performance Improvements: A Resting-State fMRI Study. *Frontiers in Human Neuroscience*, *11*(February), 1–11. http://doi.org/10.3389/fnhum.2017.00054

Robinson, D. S. (1975). Changes in monoamine oxidase and monoamines with human development and aging. In *Biology of Aging and Development* (pp. 203–212). Springer.

Rommelse, N. N. J., Van der Stigchel, S., & Sergeant, J. A. (2008). A review on eye movement studies in childhood and adolescent psychiatry. *Brain and Cognition*, *68*(3), 391–414. http://doi.org/10.1016/j.bandc.2008.08.025

Roth-Van Der Werf, T. J. M., Resing, W. C. M., & Slenders, A. P. A. C. (2002). Task similarity and transfer of an inductive reasoning training. *Contemporary Educational Psychology*, *27*(2), 296–325. http://doi.org/10.1006/ceps.2001.1096

Rothmond, D. A., Weickert, C. S., & Webster, M. J. (2012). Developmental changes in human dopamine neurotransmission: cortical receptors and terminators. *BMC Neuroscience*, *13*(1), 18.

Ryan, C. L., & Bauman, K. (2016). *Educational Attainment in the United States: 2015 Population Characteristics*. *Current Population Reports - Census*. Retrieved from https://www.census.gov/content/dam/Census/library/publications/2016/demo/p20-578.pdf

Sagi, Y., Tavor, I., Hofstetter, S., Tzur-Moryosef, S., Blumenfeld-Katzir, T., & Assaf, Y. (2012). Learning in the Fast Lane: New Insights into Neuroplasticity. *Neuron*, *73*(6), 1195–1203. http://doi.org/10.1016/j.neuron.2012.01.025

Salthouse, T. a. (2005). Relations between cognitive abilities and measures of executive functioning. *Neuropsychology*, *19*(4), 532–45. http://doi.org/10.1037/0894-4105.19.4.532

Salvucci, D. D., & Goldberg, J. H. (2000). Identifying Fixations and Saccades in Eye-Tracking Protocols. *Proceedings of the Eye Tracking Research and Applications Symposium*, 71–78. http://doi.org/10.1145/355017.355028

Samuels, E. R., & Szabadi, E. (2008). Functional neuroanatomy of the noradrenergic locus coeruleus: its roles in the regulation of arousal and autonomic function part I: principles of functional organisation. *Current Neuropharmacology*, *6*(3), 235–253. http://doi.org/10.2174/157015908785777229

Sara, S. J. (2009). The locus coeruleus and noradrenergic modulation of cognition. *Nature Reviews Neuroscience*, *10*(3), 211–223. http://doi.org/10.1038/nrn2573

Satterthwaite, T. D., Green, L., Myerson, J., Parker, J., Ramaratnam, M., & Buckner, R. L. (2007). Dissociable but inter-related systems of cognitive control and reward during decision making: Evidence from pupillometry and event-related fMRI. *NeuroImage*, *37*(3), 1017–1031. http://doi.org/10.1016/j.neuroimage.2007.04.066

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*(July), 1–10. http://doi.org/10.3389/fnagi.2010.00027

Schneider, W., Kron, V., Hünnerkopf, M., & Krajewski, K. (2004). The development of young children's memory strategies: First findings from the Würzburg Longitudinal Memory Study. *Journal of Experimental Child Psychology*, *88*(2), 193–209. http://doi.org/10.1016/j.jecp.2004.02.004

Schultz, W. (2013). Updating dopamine reward signals. *Current Opinion in Neurobiology*, *23*(2), 229–238. http://doi.org/10.1016/j.conb.2012.11.012

Schultz, W., Dayan, P., & Montague, P. R. (1997). A Neural Substrate of Prediction and Reward. *Science*, *275*(5306), 1593–1599. http://doi.org/10.1126/science.275.5306.1593

Schunk, D. H. (1996). *Learning theories*. *Printice Hall Inc., New Jersey* (Vol. 53). http://doi.org/10.1017/CBO9781107415324.004

Schweitzer, N. M. J. (1956). Threshold measurements on the light reflex of the pupil in the dark adapted eye. *Documenta Ophthalmologica*, *10*(1). http://doi.org/10.1007/BF00172098

Shepherd, M., Findlay, J. M., & Hockey, R. J. (1986). The relationship between eye movements and spatial attention. *The Quarterly Journal of Experimental Psychology Section A*, *38*(3), 475–491. http://doi.org/10.1080/14640748608401609

Shimamura, A. P., Berry, J. M., Mangels, J. A., Rusting, C. L., & Jurica, P. J. (1995). Memory and cognitive abilities in university professors: Evidence for successful aging. *Psychological Science*, *6*(5), 271–277. http://doi.org/10.1111/j.1467-9280.1995.tb00510.x

Shipstead, Z., Redick, T. S., & Engle, R. W. (2012). Is working memory training effective? *Psychological Bulletin*, *138*(4), 628–654. http://doi.org/10.1037/a0027473\r10.1037/0033-2909.131.1.30; Aiken, L.S., West, S.G., Invalidity of true experiments: Self-report pretest biases (1990) Evaluation Review, 14, pp. 374-390. , doi:10.1177/ 0193841X9001400403; Alloway, T.P., Can interactive working memory training improving learning? Journal of Interactive Learning Research, , (in press); Alloway, T.P., Alloway, R.G., Jungle memory training program (2008), Edinburgh, United Kingdom: MemosyneAlloway, T.P., Gathercole, S.E., Kirkwood, H., Elliot, J.

Shultz, S., Klin, A., & Jones, W. (2011). Inhibition of eye blinking reveals subjective perceptions of stimulus salience. *Proceedings of the National Academy of Sciences*, *108*(52), 21270–21275. http://doi.org/10.1073/pnas.1109304108

Siegle, G. J., Ichikawa, N., & Steinhauer, S. (2008). Blink before and after you think: Blinks occur prior to and following cognitive load indexed by pupillary responses. *Psychophysiology*, *45*(5), 679–687. http://doi.org/10.1111/j.1469-8986.2008.00681.x

Siegle, G. J., Steinhauer, S. R., Stenger, V. A., Konecky, R., & Carter, C. S. (2003). Use of concurrent pupil dilation assessment to inform interpretation and analysis of fMRI data. *NeuroImage*, *20*(1), 114–124. http://doi.org/10.1016/S1053-8119(03)00298-2

Siegle, G. J., Steinhauer, S. R., & Thase, M. E. (2004). Pupillary assessment and computational modeling of the Stroop task in depression. *International Journal of Psychophysiology*, *52*(1), 63–76. http://doi.org/10.1016/j.ijpsycho.2003.12.010

Simon, N. W., Gregory, T. A., Wood, J., & Moghaddam, B. (2013). Differences in response initiation and behavioral flexibility between adolescent and adult rats. *Behavioral Neuroscience*, *127*(1), 23–32. http://doi.org/10.1037/a0031328

Singer, H. S., Butler, I. J., Tune, L. E., Seifert, W. E., & Coyle, J. T. (1982). Dopaminergic dysfunction in Tourette syndrome. *Annals of Neurology*, *12*(4), 361–366.

Sirois, S., & Brisson, J. (2014). Pupillometry. *Wiley Interdisciplinary Reviews: Cognitive Science*, *5*(6). http://doi.org/10.1002/wcs.1323

Sirois, S., & Jackson, I. R. (2012). Pupil Dilation and Object Permanence in Infants. *Infancy*, *17*(1), 61–78. http://doi.org/10.1111/j.1532-7078.2011.00096.x

Slagter, H. a, Georgopoulou, K., & Frank, M. J. (2015). Spontaneous eye blink rate predicts learning from negative, but not positive, outcomes. *Neuropsychologia*, *71*, 126–132. http://doi.org/10.1016/j.neuropsychologia.2015.03.028

Smilek, D., Carriere, J. S. A., & Cheyne, J. A. (2010). Out of Mind, Out of Sight: Eye Blinking as Indicator and Embodiment of Mind Wandering. *Psychological Science*, *21*(6), 786–789. http://doi.org/10.1177/0956797610368063

Smith, L. B., Yu, C., Yoshida, H., & Fausey, C. M. (2015). Contributions of Head-Mounted Cameras to Studying the Visual Environments of Infants and Young Children. *Journal of Cognition and Development*, *16*(3), 407–419. http://doi.org/10.1080/15248372.2014.933430

Smith, S. M., Fox, P. T., Miller, K. L., Glahn, D. C., Fox, P. M., Mackay, C. E., … Beckmann, C. F. (2009). Correspondence of the brain's functional architecture during activation and rest. *Proceedings of the National Academy of Sciences*, *106*(31), 13040–13045. http://doi.org/10.1073/pnas.0905267106

Soares, J. M., Sampaio, A., Ferreira, L. M., Santos, N. C., Marques, P., Marques, F., … Sousa, N. (2013). Stress Impact on Resting State Brain Networks. *PLoS ONE*, *8*(6), e66500. http://doi.org/10.1371/journal.pone.0066500

Söderqvist, S., Bergman Nutley, S., Peyrard-Janvid, M., Matsson, H., Humphreys, K., Kere, J., & Klingberg, T. (2012). Dopamine, working memory, and training induced plasticity: implications for developmental research. *Developmental Psychology*, *48*(3), 836–43. http://doi.org/10.1037/a0026179

Steele, C. M. (1997). A threat in the air. How stereotypes shape intellectual identity and performance. *The American Psychologist*, *52*(6), 613–629. http://doi.org/10.1037/0003-066X.52.6.613

Steinhauer, S. R., Siegle, G. J., Condray, R., & Pless, M. (2004). Sympathetic and parasympathetic innervation of pupillary dilation during sustained processing. *International Journal of Psychophysiology*, *52*(1), 77–86. http://doi.org/10.1016/j.ijpsycho.2003.12.005

Stern, J. A., Boyer, D., & Schroeder, D. (1994). Blink rate: A possible measure of fatigue. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *36*(2), 285–297.

Sutton, S., Braren, M., Zubin, J., & John, E. R. (1965). Evoked-potential correlates of stimulus uncertainty. *Science (New York, N.Y.)*, *150*(3700), 1187–1188.

Takahashi, H. (2013). PET neuroimaging of extrastriatal dopamine receptors and prefrontal cortex functions. *Journal of Physiology Paris*, *107*(6), 503–509. http://doi.org/10.1016/j.jphysparis.2013.07.001

Takeuchi, H., Taki, Y., Nouchi, R., Hashizume, H., Sekiguchi, A., Kotozaki, Y., … Kawashima, R. (2013). Effects of working memory training on functional connectivity and cerebral blood flow during rest. *Cortex*, *49*(8), 2106–2125. http://doi.org/10.1016/j.cortex.2012.09.007

Taubert, M., Lohmann, G., Margulies, D. S., Villringer, A., & Ragert, P. (2011). Long-term effects of motor training on resting-state networks and underlying brain structure. *Neuroimage*, *57*(4), 1492–1498. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=21672633

Taylor, J. R., Elsworth, J. D., Lawrence, M. S., Sladek, J. R., Roth, R. H., & Redmond, D. E.

(1999). Spontaneous blink rates correlate with dopamine levels in the caudate nucleus of MPTP-treated monkeys. *Experimental Neurology*, *158*, 214–220. http://doi.org/10.1006/exnr.1999.7093

Taylor, V. a, Daneault, V., Grant, J., Scavone, G., Breton, E., Roffe-Vidal, S., … Beauregard, M. (2013). Impact of meditation training on the default mode network during a restful state. *Social Cognitive and Affective Neuroscience*, *8*(1), 4–14. http://doi.org/10.1093/scan/nsr087

Tenenbaum, E. J., Shah, R. J., Sobel, D. M., Malle, B. F., & Morgan, J. L. (2013). Increased Focus on the Mouth Among Infants in the First Year of Life: A Longitudinal Eye-Tracking Study. *Infancy*, *18*(4), 534–553. http://doi.org/10.1111/j.1532-7078.2012.00135.x

Tharp, I. J., & Pickering, A. D. (2011). Individual differences in cognitive-flexibility: The influence of spontaneous eyeblink rate, trait psychoticism and working memory on attentional set-shifting. *Brain and Cognition*, *75*(2), 119–125. http://doi.org/10.1016/j.bandc.2010.10.010

Tharp, J. A., Wendelken, C., Mathews, C. A., Marco, E. J., Schreier, H., & Bunge, S. A. (2015). Tourette syndrome: Complementary insights from measures of cognitive control, eyeblink rate, and pupil diameter. *Frontiers in Psychiatry*, *6*(JUN), 1–10. http://doi.org/10.3389/fpsyt.2015.00095

Theeuwes, J., Belopolsky, A., & Olivers, C. N. L. (2009). Interactions between working memory, attention and eye movements. *Acta Psychologica*, *132*(2), 106–114. http://doi.org/10.1016/j.actpsy.2009.01.005

Thibaut, J.-P., French, R. M., Missault, A., Gérard, Y., & Glady, Y. (2011). In the Eyes of the Beholder: What Eye-Tracking Reveals About Analogy-Making Strategies in Children and Adults. In *Proceedings of the thirty-third Annual Meeting of the Cognitive Science Society*.

Thibaut, J., & French, R. (2011). In the eyes of the beholder: What eye-tracking reveals about analogy-making strategies in children and adults. *Proceedings of the …*, 453–458. Retrieved from http://palm.mindmodeling.org/cogsci2011/papers/0085/paper0085.pdf

Thibaut, J. P., & French, R. M. (2016). Analogical reasoning, control and executive functions: A developmental investigation with eye-tracking. *Cognitive Development*, *38*, 10–26. http://doi.org/10.1016/j.cogdev.2015.12.002

Thomas, L. E., & Lleras, A. (2007). Moving eyes and moving thought: on the spatial compatibility between eye movements and cognition. *Psychonomic Bulletin & Review*, *14*(4), 663–668.

Tiadi, A., Gérard, C.-L., Peyre, H., Bui-Quoc, E., & Bucci, M. P. (2016). Immaturity of Visual Fixations in Dyslexic Children. *Frontiers in Human Neuroscience*, *10*. http://doi.org/10.3389/fnhum.2016.00058

Tidwell, J. W., Dougherty, M. R., Chrabaszcz, J. R., Thomas, R. P., & Mendoza, J. L. (2014). What counts as evidence for working memory training? Problems with correlated gains and dichotomization. *Psychonomic Bulletin & Review*, *21*(3), 620–8.

http://doi.org/10.3758/s13423-013-0560-7

Traxler, M. J., Williams, R. S., Blozis, S. A., & Morris, R. K. (2005). Working memory, animacy, and verb class in the processing of relative clauses. *Journal of Memory and Language*, *53*(2), 204–224. http://doi.org/10.1016/j.jml.2005.02.010

Tulen, J. H. M., Azzolini, M., De Vries, J. A., Groeneveld, W. H., Passchier, J., & Van De Wetering, B. J. M. (1999). Quantitative study of spontaneous eye blinks and eye tics in Gilles de la Tourette's syndrome. *Journal of Neurology, Neurosurgery & Psychiatry*, *67*(6), 800–802.

Urner, M., Schwarzkopf, D. S., Friston, K., & Rees, G. (2013). Early visual learning induces long-lasting connectivity changes during rest in the human brain. *NeuroImage*, *77*, 148–56. http://doi.org/10.1016/j.neuroimage.2013.03.050

Vahdat, S., Darainy, M., Milner, T. E., & Ostry, D. J. (2011). Functionally Specific Changes in Resting-State Sensorimotor Networks after Motor Learning. *Journal of Neuroscience*, *31*(47), 16907–16915. http://doi.org/10.1523/JNEUROSCI.2737-11.2011

Vaisvaser, S., Lin, T., Admon, R., Podlipsky, I., Greenman, Y., Stern, N., … Hendler, T. (2013). Neural traces of stress: cortisol related sustained enhancement of amygdala-hippocampal functional connectivity. *Frontiers in Human Neuroscience*, *7*(July), 1–11. http://doi.org/10.3389/fnhum.2013.00313

van Bochove, M. E., Van der Haegen, L., Notebaert, W., & Verguts, T. (2013). Blinking predicts enhanced cognitive control. *Cognitive, Affective & Behavioral Neuroscience*, *13*(2), 346–54. http://doi.org/10.3758/s13415-012-0138-2

van der Lans, R., Wedel, M., & Pieters, R. (2011). Defining eye-fixation sequences across individuals and tasks: the Binocular-Individual Threshold (BIT) algorithm. *Behavior Research Methods*, *43*(1), 239–257. http://doi.org/10.3758/s13428-010-0031-2

van der Post, J., de Waal, P. P., de Kam, M. L., Cohen, A. F., & van Gerven, J. M. A. (2004). No evidence of the usefulness of eye blinking as a marker for central dopaminergic activity. *Journal of Psychopharmacology*, *18*(1), 109–114. http://doi.org/10.1177/0269881104042832

van der Schaaf, M. E., van Schouwenburg, M. R., Geurts, D. E. M., Schellekens, A. F. A., Buitelaar, J. K., Verkes, R. J., & Cools, R. (2014). Establishing the Dopamine Dependency of Human Striatal Signals during Reward and Punishment Reversal Learning. *Cerebral Cortex*, *24*(3), 633–642. http://doi.org/10.1093/cercor/bhs344

Van der Stigchel, S., Meeter, M., & Theeuwes, J. (2006). Eye movement trajectories and what they tell us. *Neuroscience & Biobehavioral Reviews*, *30*(5), 666–679. http://doi.org/10.1016/j.neubiorev.2005.12.001

van Hell, H. H., Vink, M., Ossewaarde, L., Jager, G., Kahn, R. S., & Ramsey, N. F. (2010). Chronic effects of cannabis use on the human reward system: An fMRI study. *European Neuropsychopharmacology*, *20*(3), 153–163.

http://doi.org/10.1016/j.euroneuro.2009.11.010

Varazzani, C., San-Galli, A., Gilardeau, S., & Bouret, S. (2015). Noradrenaline and Dopamine Neurons in the Reward/Effort Trade-Off: A Direct Electrophysiological Comparison in Behaving Monkeys. *Journal of Neuroscience*, *35*(20), 7866–7877. http://doi.org/10.1523/JNEUROSCI.0454-15.2015

Vendetti, M. S., & Bunge, S. a. (2014). Perspective Evolutionary and Developmental Changes in the Lateral Frontoparietal Network : A Little Goes a Long Way for Higher-Level Cognition. *Neuron*, *84*(5), 906–917. http://doi.org/10.1016/j.neuron.2014.09.035

Vendetti, M. S., Matlen, B. J., Richland, L., & Bunge, S. A. (2015). Analogical reasoning in the classroom: Insights from cognitive science. *Mind, Brain, and Education*, *9*(2). http://doi.org/10.1016/j.socscimed.2013.05.029

Verney, S. P., Granholm, E., & Marshall, S. P. (2004). Pupillary responses on the visual backward masking task reflect general cognitive ability. *International Journal of Psychophysiology*, *52*(1), 23–36. http://doi.org/10.1016/j.ijpsycho.2003.12.003

Vigneau, F., Caissie, A. F., & Bors, D. A. (2006). Eye-movement analysis demonstrates strategic influences on intelligence. *Intelligence*, *34*(3), 261–272. http://doi.org/10.1016/j.intell.2005.11.003

Wade, N. J. (2015). How Were Eye Movements Recorded Before Yarbus? *Perception*, *44*(8–9), 851–83. http://doi.org/10.1177/0301006615594947

Wagenmakers, E. J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., … Morey, R. D. (2017). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin and Review*, 1–19. http://doi.org/10.3758/s13423-017-1323-7

Walberg, H. J., & Tsai, S.-L. (1983). Matthew Effects in Education. *American Educational Research Journal*, *20*(3), 359–373. http://doi.org/10.3102/00028312020003359

Walker, M. P., & Stickgold, R. (2006). Sleep, Memory, and Plasticity. *Annual Review of Psychology*, *57*(1), 139–166. http://doi.org/10.1146/annurev.psych.56.091103.070307

Wang, Z., Zhou, R., & Shah, P. (2014). Spaced cognitive training promotes training transfer. *Frontiers in Human Neuroscience*, *8*(April), 1–8. http://doi.org/10.3389/fnhum.2014.00217

Waterhouse, B. D., & Woodward, D. J. (1980). Interaction of norepinephrine with cerebrocortical activity evoked by stimulation of somatosensory afferent pathways in the rat. *Experimental Neurology*, *67*(1), 11–34. http://doi.org/10.1016/0014-4886(80)90159-4

Weickert, C. S., Webster, M. J., Gondipalli, P., Rothmond, D., Fatula, R. J., Herman, M. M., … Akil, M. (2007). Postnatal alterations in dopaminergic markers in the human prefrontal cortex. *Neuroscience*, *144*(3), 1109–19. http://doi.org/10.1016/j.neuroscience.2006.10.009

Wendelken, C., & Bunge, S. A. (2009a). Transitive Inference: Distinct Contributions of Rostrolateral Prefrontal Cortex and the Hippocampus. *Journal of Cognitive Neuroscience*, 1–11.

Wendelken, C., & Bunge, S. a. (2009b). Transitive Inference: Distinct Contributions of Rostrolateral Prefrontal Cortex and the Hippocampus. *Journal of Cognitive Neuroscience*, 1–11.

Wendelken, C., Ferrer, E., Whitaker, K. J., & Bunge, S. A. (2016). Fronto-Parietal Network Reconfiguration Supports the Development of Reasoning Ability. *Cerebral Cortex*, *26*(5), 2178–2190. http://doi.org/10.1093/cercor/bhv050

Werchan, D. M., Collins, A. G. E., Frank, M. J., & Amso, D. (2015). 8-Month-Old Infants Spontaneously Learn and Generalize Hierarchical Rules. *Psychological Science*, *26*(6), 805–815. http://doi.org/10.1177/0956797615571442

Westbrook, A., & Braver, T. S. (2016). Dopamine Does Double Duty in Motivating Cognitive Effort. *Neuron*, *89*(4), 695–710. http://doi.org/10.1016/j.neuron.2015.12.029

Wetzel, N., Buttelmann, D., Schieler, A., & Widmann, A. (2016). Infant and adult pupil dilation in response to unexpected sounds. *Developmental Psychobiology*, *58*(3), 382–392. http://doi.org/10.1002/dev.21377

Wierda, S. M., van Rijn, H., Taatgen, N. A., & Martens, S. (2012). Pupil dilation deconvolution reveals the dynamics of attention at high temporal resolution. *Proceedings of the National Academy of Sciences*, *109*(22), 8456–8460. http://doi.org/10.1073/pnas.1201858109

Wiseman, R. J., & Nakano, T. (2016). Blink and you'll miss it: the role of blinking in the perception of magic tricks. *PeerJ*, *4*, e1873. http://doi.org/10.7717/peerj.1873

Woodcock, R. W., McGrew, K., & Mather, N. (2001). *Woodcock-Johnson III Tests of Cognitive Abilities*. (N. JA & S. Goldstein, Eds.). Riverside Publishing.

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), 247–261. http://doi.org/10.1037/h0074898

Yarbus, A. L. (1967). *Eye movements during perception of complex objects*. New York: Plenum Press.

Yellin, D., Berkovich-Ohana, A., & Malach, R. (2015). Coupling between pupil fluctuations and resting-state fMRI uncovers a slow build-up of antagonistic responses in the human cortex. *NeuroImage*, *106*, 414–427. http://doi.org/10.1016/j.neuroimage.2014.11.034

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, *18*(5), 459–482. http://doi.org/10.1002/cne.920180503

Yin, H. H., & Knowlton, B. J. (2006). The role of the basal ganglia in habit formation. *Nature Reviews. Neuroscience*, *7*(6), 464–76. http://doi.org/10.1038/nrn1919

Yoo, K., Sohn, W. S., & Jeong, Y. (2013). Tool-use practice induces changes in intrinsic functional connectivity of parietal areas. *Frontiers in Human Neuroscience*, *7*(February), 1–9. http://doi.org/10.3389/fnhum.2013.00049

Yoon, D., & Narayanan, N. H. (2004). Mental imagery in problem solving: An eye tracking study. *The 2004 Symposium on Eye Tracking Research & Applications*, *1*(212), 77–84. http://doi.org/http://doi.acm.org/10.1145/968363.968382

Yu, A., & Dayan, P. (2003). Expected and unexpected uncertainty: ACh and NE in the neocortex. *Advances in Neural Information Processing Systems*, 173–180.

Yu, A. J., & Dayan, P. (2005). Uncertainty, neuromodulation, and attention. *Neuron*, *46*(4), 681–692. http://doi.org/10.1016/j.neuron.2005.04.026

Yu, C., & Smith, L. B. (n.d.). From infant hands to parent eyes: Hand-eye coordination predicts joint attention.

Yu, C., & Smith, L. B. (2016). The Social Origins of Sustained Attention in One-Year-Old Human Infants. *Current Biology*, *26*(9), 1235–1240. http://doi.org/10.1016/j.cub.2016.03.026

Zaman, M. L., & Doughty, M. J. (1997). Some methodological issues in the assessment of the spontaneous eyeblink frequency in man. *Ophthalmic and Physiological Optics*, *17*(5), 421–432.

Zametkin, A. J., Stevens, J. R., & Pittman, R. (1979). Ontogeny of spontaneous blinking and of habituation of the blink reflex. *Annals of Neurology*, *5*(5), 453–457. http://doi.org/10.1002/ana.410050509

Zatorre, R. J., Fields, R. D., & Johansen-Berg, H. (2012). Plasticity in gray and white: Neuroimaging changes in brain structure during learning. *Nature Neuroscience*, *15*(4), 528–536. http://doi.org/10.1038/nn.3045

Zhang, T., Mou, D., Wang, C., Tan, F., Jiang, Y., Lijun, Z., & Li, H. (2015). Dopamine and executive function: Increased spontaneous eye blink rates correlate with better set-shifting and inhibition, but poorer updating. *International Journal of Psychophysiology*, *96*(3), 155–161. http://doi.org/10.1016/j.ijpsycho.2015.04.010

# Appendix A

## Supplementary information for Chapter 3

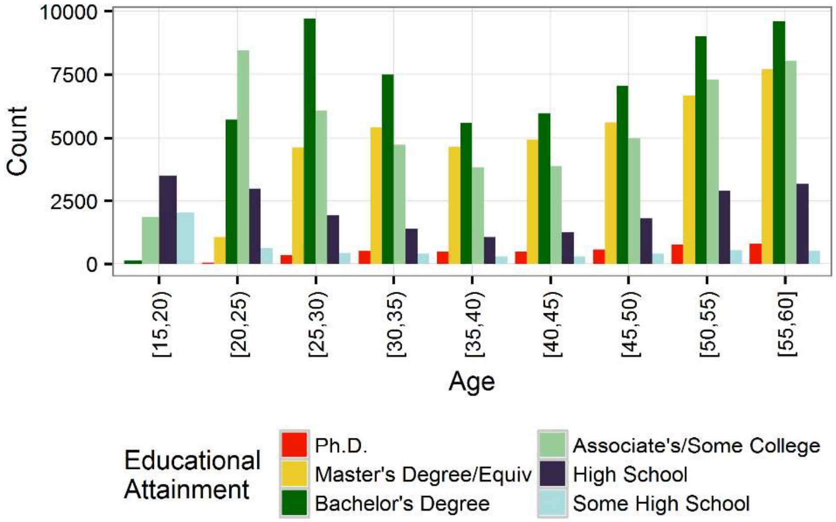


**Figure S1. Distribution of educational attainment across age groups.** The number of participants between the ages of 15-60, reporting educational attainment between Some High School to Ph.D.'s at T1 (*N* = 196,388).

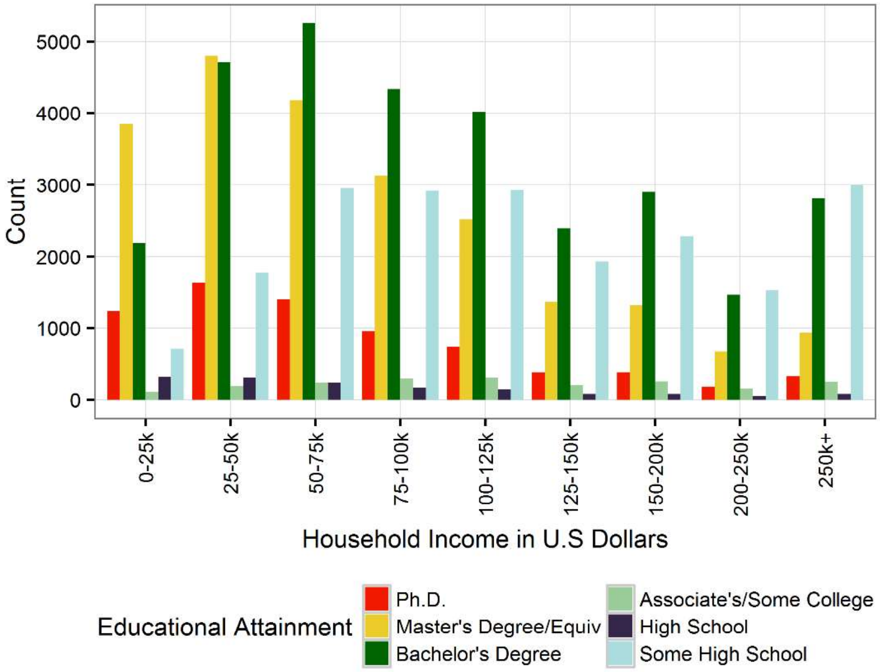


**Figure S2. Distribution of educational attainment across household income.** The number of participants who identified their household income bracket in U.S. dollars, and reported their educational attainment between Some High School to Ph.D.'s at T1 (*N* = 196,388).

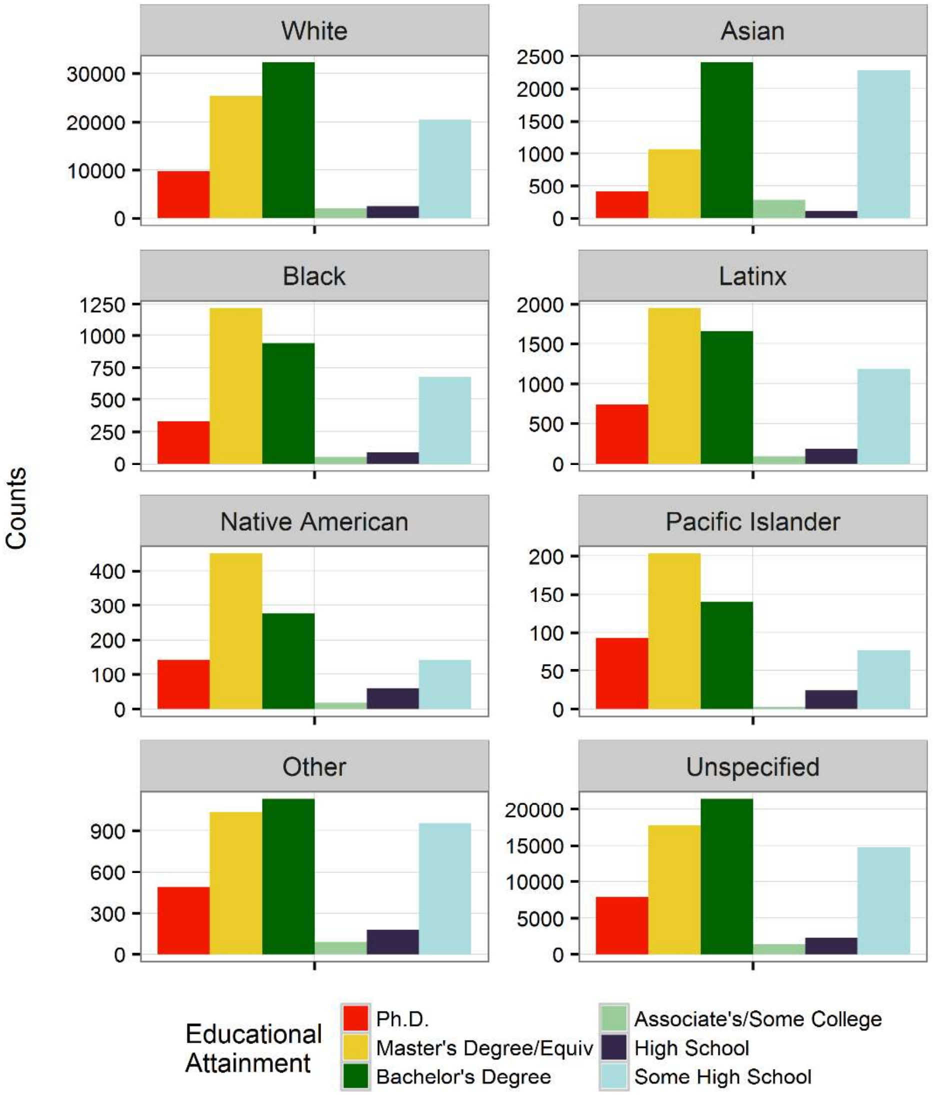


**Figure S3. Distribution of educational attainment across ethnicity categories.** The number of participants across ethnic categories, reporting educational attainment between Some High School to Ph.D.'s at T1 (*N* = 196,388).

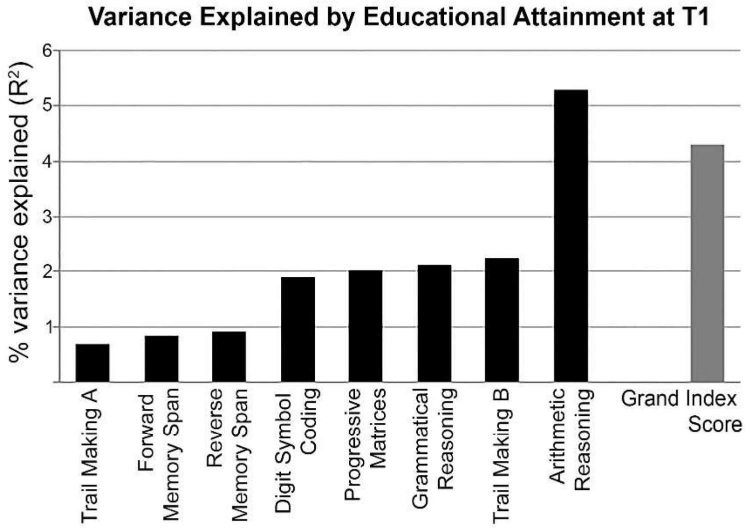


**Figure S4. Unique variance in cognitive performance explained by educational attainment at T1.** Displaying $R^2$ values from regression models of educational attainment predicting performance on individual cognitive assessments.

**Table S1**. Demographic and engagement covariates predicting T1 cognitive performance and learning, as measured by the Grand Index score (GI)

| **Covariate Type** | | **GI T1 Estimate** | **Δ GI Estimate** |
|---|---|---|---|
| | Intercept | 100.69*** | 6.83*** |
| **Engagement** | Lumosity game play time (h) | 2.67*** | 1.37*** |
| | Days between T1 and T2 | | -2.51*** |
| **Native language** | English not native language | -5.77*** | -0.52*** |
| | Native language unspecified | -0.65** | -0.52** |
| **Gender** | Female | -3.79*** | -0.80*** |
| | Gender unspecified | -3.60*** | -0.23 |
| **Ethnicity** | Asian | 5.63*** | 0.43** |
| | Black | -7.08*** | -1.26*** |
| | Latinx | -1.98*** | -0.57*** |
| | Native American | -2.16*** | -0.36 |
| | Pacific Islander | -0.21 | 0.58 |
| | Other | -2.88*** | -0.39* |
| | Unspecified | -1.18*** | 0.10 |
| **Household Income** | 0-25k | 0.20 | 0.23 |
| | 25-50k | 0.16 | 0.09 |
| | 50-75k | 0.25 | 0.09 |
| | 100-125k | 0.22 | 0.15 |
| | 125-150k | 0.70** | 0.15 |
| | 150-200k | 0.97*** | 0.06 |
| | 200-250k | 0.77** | -0.01 |
| | Over 250k | 0.83*** | 0.38* |
| | Non US income | 0.74*** | 0.50** |
| | Unspecified | -1.01*** | 0.05 |
| **T1 Performance** | | | -0.10*** |

Together, all these variables account for 5% of the variance in cognitive performance at both timepoints. It is plausible that the race effects are confounded by childhood SES, which we did not measure. The GI change score (ΔGI) is the difference between the GI score from T1 and T2. Each GI score was normalized to have a distribution with mean of 100 (15 SD) Reference categories: English speaker, Male, White, 75-100k household income. The logarithmic form of the engagement variables was used in the analyses. $p < 0.0001$‘***’, $p < 0.001$‘**’, $p < 0.01$‘*’

# Appendix B

## Supplementary information for Chapter 5

### Sample

**Table S1.** Participant Information

| | **Reasoning group** | **Active control** | |
|---|---|---|---|
| | Logic Games Course<br>*n* = 23 | Comprehension Course<br>*n* = 24 | Pairwise comparisons<br>*t or* $X^2$ |
| Females | 14 | 13 | *0.12* |
| Age | 21.55 (3.96) | 21.88 (4.93) | 0.25 |
| WJ III Analysis Synthesis | 29.39 (1.90) | 30.43 (2.61) | 1.55 |
| Digit Span | 7.00 (2.34) | 6.92 (1.18) | -0.16 |
| Spatial Span | 6.39 (1.53) | 6.38 (0.97) | -0.04 |
| Logic Games sample problems | 1.74 (1.25) | 1.58 (1.10) | -0.45 |
| Reading Comprehension sample problems | 2.61 (1.28) | 2.62 (1.12) | 0.05 |
| LSAT score (T1 practice exam) | 145 | 147 | -0.83 |

Mean and SDs. *t* or $X^2$ values reported for the variables the groups were matched on at T1 (p's > 0.54, Analysis Synthesis p=0.14). Performance on assessments as raw scores. The maximum score on the sample LSAT Logic Games and Reading Comprehension problems are six points. LSAT scores are based on the full practice exam administered at T1 and on the standard 120-180 LSAT point scale.

## LSAT manipulation

### S1. Improvements in LSAT performance

We tested for changes in performance on a set of Logic Games and Reading Comprehension problems administered in the lab at each timepoint[3]. Each set consisted of 6 problems. An additional set of 6 questions per section was administered at post-test[4]. Participants had up to eight minutes to answer each problem set.We found no significant effect of Time, Group, or their interaction on performance on the Reading Comprehension problems (all $p > 0.36$) or the Logic Games problems (all $p > 0.14$) administered at both timepoints. On the new set of problems administered only at post-test, the Comprehension group performed better on the Reading Comprehension problems ($t(42) = 2.07$, $p = 0.05$), and incorporated the strategies taught in their course more when solving these problems (Table S2). However, the groups did not differ in performance on the Logic Games problems ($t(42) = -0.66$, $p = 0.51$). Although participants in both groups reported after the post-test assessment that they thought they had improved most on the section they had studied (Table S2), it was not the case that the Logic Games group improved more than the other group on Logic Games – at least, not on the basis of the limited number of LSAT problems administered in the laboratory.

**Table S2.** LSAT self-report data and strategy use ratings

| | **Reasoning group** Logic Games Course $n = 23$ | **Active control** Comprehension Course $n = 24$ |
|---|---|---|
| Rating of course effectiveness | 4 ("effective") | 4 ("effective") |
| Rating of course enjoyment | 3 ("neutral") | 3 ("neutral") |
| Perceived improvements on Logic Games | 4 ("stronger performance") | 3 ("no change") |
| Perceived improvements on Reading Comprehension | 3 ("no change") | 4 ("stronger performance") |
| % change in Logic Games strategy-use | 11% | 11% |
| % change in Reading Comprehension strategy-use | 25% | 56% |

Median of self-reported enjoyment, course effectiveness, and perceived improvement on a 5-point Likert scale, showing in parenthesis the answer choice equivalent to the numerical rating. The original scale of "perceived improvements" was recoded from a range of [-2, +2] to [1, 5] on this table to be consistent with the other measures. Use of LSAT strategies could amount to 10 points each. Showing the percent change for each section, calculated using the median score at each timepoint. Scoring of strategies available upon request.

---

[3] Game 3.3, LSAT 02/1992 Prep Test 4; Passage 4.2, LSAT 10/2011 Prep Test 64

[4] Game 2.4, LSAT 02/1994 Prep Test 10; Passage 3.2, LSAT 10/1994 Prep Test 12

## Transitive inference task: gaze and behavioral metrics

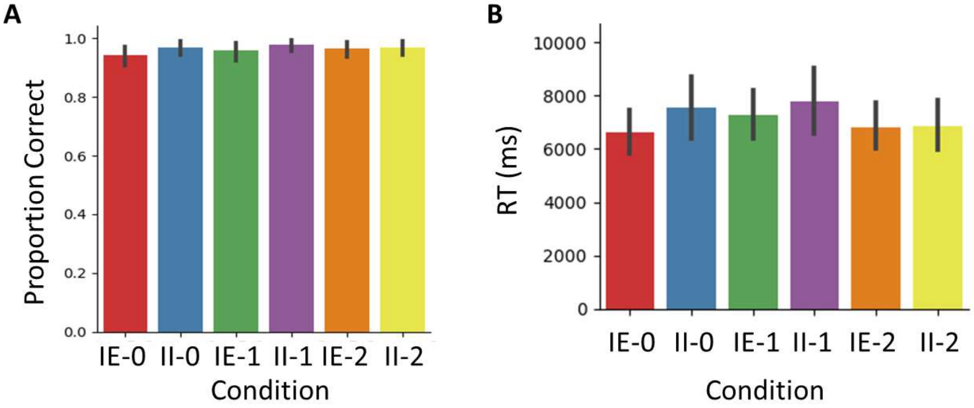


**Fig S1. Effect of problem type on behavioral performance on the transitive inference task at pretest.** The order in which relevant relations appeared (i.e., 0-2 scales apart) or whether the question involved integrating between inequalities (II) or between an inequality and an equality (IE) did not significantly impact either a) proportion of correct responses, or b) RTs on accurate trials. All error bars are 95% C.I estimated with 5000 bootstrap iterations

## S2. Gaze data preprocessing and fixation detection algorithm

We used custom scripts written in Python v3.6 to preprocess the gaze data and identify fixations. We classified gaze data points into fixations using a dispersion-based algorithm (Salvucci & Goldberg, 2000), such that in a 100ms window the gaze position could not exceed a radius of 35 pixels (1° visual angle) to be considered part of the same fixation. That distance has been previously identified as a reliable threshold for accurate fixation classification (van der Lans, Wedel, & Pieters, 2011) and is sufficient given the size of our stimuli. Single gaze points that deviated from the distance threshold were considered part of the same ongoing fixation as long as the next legitimate gaze measure returned to the accepted threshold. Gaps between gaze data points of less than 40ms were ignored, as it is physiologically unlikely that a separate fixation or blink occurred during that time; it is more likely that those gaps represent an artifact of the apparatus (Holmqvist et al., 2011). Both groups showed a reduction in total fixation times at T2 ($t(44) = -3.52$, $p < 0.01$), but the groups did not significantly differ on total fixation times at either timepoint ($p > 0.2$).

**S3. Rationale behind gaze metric development**. We had initially theorized that reasoning would proceed in three discrete stages: first, identification of the relevant scales (visual search); then, a singular focus on these relevant scales, aimed at encoding the relations they depicted (relational encoding); and finally, joint consideration of these relations as individuals engaged in the last step of transitive inference (relational integration). However, an analysis of sequences of fixations, collapsing across groups and timepoints, provided two clues that the process of reasoning was not strictly stage-like, as described below.

First, fixations on irrelevant scales did not cease abruptly after an initial search of the array; rather, they tapered off slowly over the course of the trial (Fig S2). As such, rather than measuring visual search as the number of fixations a participant made before they stopped looking at the irrelevant scales anymore, we identified the point in the trial at which the probability of looking at an irrelevant scale dipped below chance, and the probability of looking at a relevant scale rose above chance.

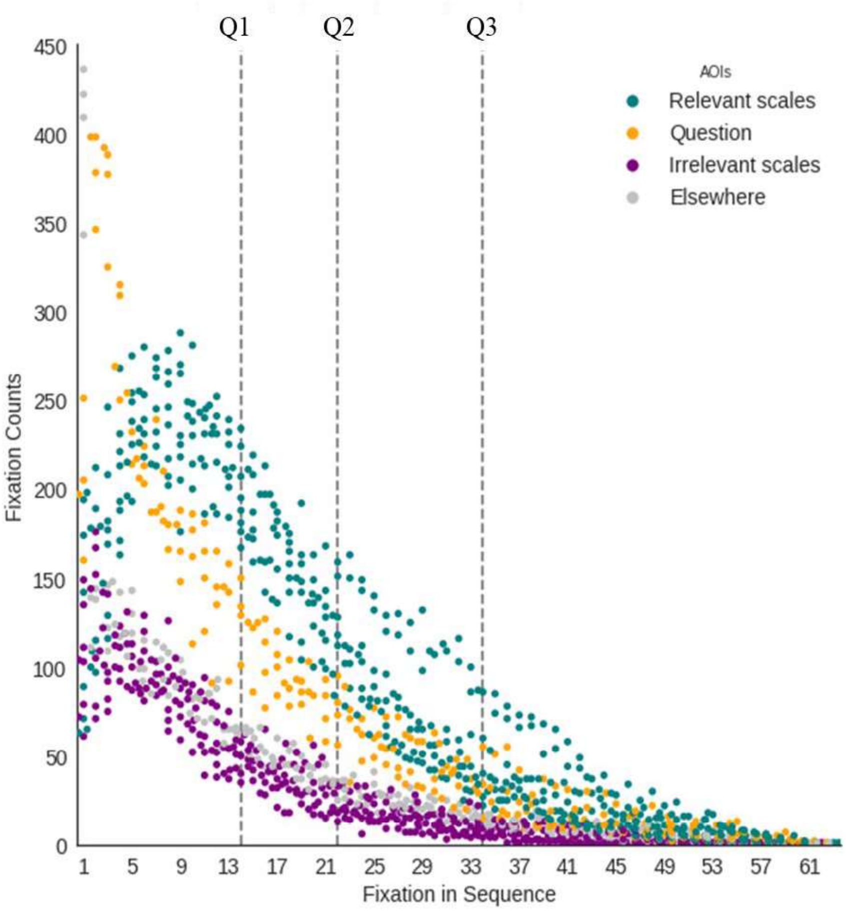


**Fig S2. Fixations patterns on the transitive inference task.** Plotting fixations during problem solving across groups and timepoints. Participants made a median of 22 fixations on correct problems. Trials with up to 64 fixations were included in the analyses (i.e., range x-axis). Vertical dotted lines denote quartiles of total fixations (e.g., the vertical line denoting Q1, the first quartile, indicates that 75% of the trials had at least 14 fixations. Only 25% of trials had more than 34 fixations, as denoted by the Q3 line). Colored dots represent the total number of fixations (y-axis) across both groups and timepoints. The colors indicate the areas of interest (AOIs) where those fixations occurred: two relevant scales (teal), two irrelevant scales (purple), and the question area (yellow). In gray, fixations not on an AOI, but elsewhere on the screen.

(S2 continues) Second, the duration of fixations on relevant scales did not decrease over the course of a trial, as it should if this were a pure metric of relational encoding; rather, it increased (Fig. S3). This finding intimates that long fixations on a relevant scale towards the end of the trial reflect simultaneous consideration of that scale and the other relevant scale – i.e., relational integration. As such, we renamed the metric that we had previously labeled ‘relational encoding’ to ‘relational thinking’, to denote the fact that it likely reflects relational encoding and maintenance towards the beginning of the trial (after preferentially looking at the relevant scales), and relational integration towards the end of the trial.

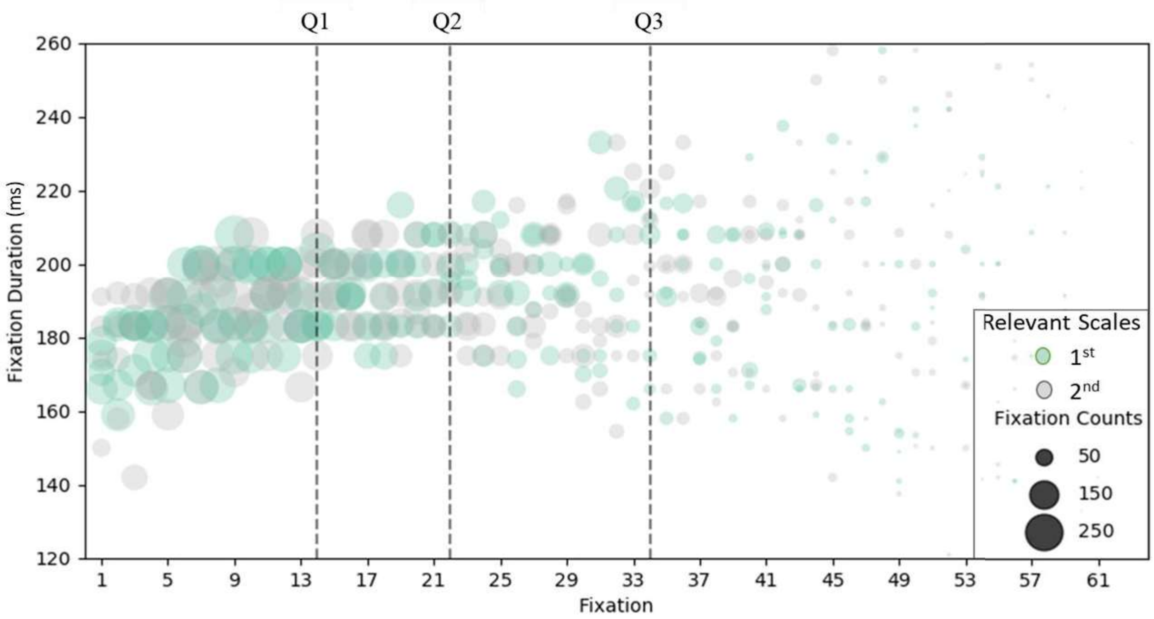


**Fig S3. Count and median duration of fixations on relevant scales on the transitive inference task.** Plotting fixations on relevant scales across groups and timepoints on accurate problems. Trials up to 64 fixations were included in the analyses (i.e., range in x-axis). Vertical dotted lines denote quartiles of total fixations. Median fixation duration (y-axis) and total fixation count (size of circles) calculated across groups and timepoints. The colors indicate whether fixations occurred on the first (i.e., most-left during a trial; green) or second (gray) relevant scale.

## Bayesian Analysis

**Table S3**. Posterior odds and model comparison: gaze metrics and composite reasoning score

| **Metric** | **Models** | **P(M)** | **P(M\|data)** | **$BF_M$** | **$BF_{10}$** | **error %** |
|---|---|---|---|---|---|---|
| Visual search | Null model (incl. Group, Time, Subject) | 0.50 | 0.63 | 1.73 | 1.00 | |
| | Group x Time | 0.50 | 0.37 | 0.60 | 0.60 | 2.79 |
| Relational Thinking | Null model (incl. Group, Time, Subject) | 0.50 | 0.21 | 0.27 | 1.00 | |
| | Group x Time | 0.50 | 0.79 | 3.66 | 3.66 | 1.98 |
| Relational Integration | Null model (incl. Group, Time, Subject) | 0.50 | 0.40 | 0.68 | 1.00 | |
| | Group x Time | 0.50 | 0.60 | 1.58 | 1.58 | 2.62 |
| Composite Reasoning | Null model (incl. Group, Time, Subject) | 0.50 | 0.06 | 0.07 | 1.00 | |
| | Group x Time | 0.50 | 0.94 | 15.06 | 15.06 | 3.04 |

Estimations made using BayesFactor's (Morey et al., 2015) default Cauchy prior scale $r = \frac{1}{2}\sqrt{2}$

**Table S4**. Performance on the behavioral assessments

| Composite Measure | Subtest | Description | Evidence supporting $H_1$<br>$BF_{10} \approx P(H_1 \mid \text{data}) / P(H_0 \mid \text{data})$<br>$H_1 = POST < PRE$ | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | Reasoning | | | Comprehension | | |
| | | | PRE *M* (± *S.D*) | POST *M* (± *S.D*) | *$BF_{10}$* *(%error)* | PRE *M* (± *S.D*) | POST *M* (± *S.D*) | *$BF_{10}$* *(%error)* |
| Planning | Spatial Slider<br>Hampshire Tree Task | Reconfigure numbered shapes in ascending order using as few moves as possible | 0.43 (0.19) | 0.46 (0.20) | $0.25^{\bullet}$ (±<0.00) | 0.49 (0.20) | 0.55 (0.23) | 0.50 (±<0.00) |
| Working Memory | Digit Span Forward<br>Spatial Span Backward | Hold in mind strings of digits or spatial locations and recall them in the order in which they had appeared | 0.53 (0.16) | 0.55 (0.17) | $0.20^{\bullet}$ (±<0.00) | 0.59 (0.21) | 0.56 (0.20) | $0.34^{\bullet}$ (±0.01) |
| Selective Attention | Feature Match | Decide whether all the geometrical elements in a visual array were identical | 0.52 (0.24) | 0.57 (0.24) | 0.6 (±0.01) | 0.56 (0.26) | 0.64 (0.24) | 0.99 (±<0.00) |
| Verbal Fluency | Grammatical Reasoning | Make speeded judgments as to whether a statement correctly describes a pair of objects | 0.38 (0.20) | 0.33 (0.25) | $0.19^{\bullet}$ (±<0.00) | 0.38 (0.25) | 0.41 (0.30) | $0.37^{\bullet}$ (±<0.00) |

*$H_1 = POST < PRE$* assessed with Bayesian single-sided paired t-test. Interaction models tested with Bayesian mixed regressions. Estimations made using BayesFactor's (Morey et al., 2015) default Cauchy prior scale $r = \frac{1}{2}\sqrt{2}$ and prior uniform probability to the models. Approximate classification scheme for the interpretation of Bayes factors from (Wagenmakers et al., 2017): *Moderate evidence for *$H_1$*,, $^{\bullet}$ Moderate evidence for *$H_0$.*