Change Detection as a Framework for Understanding Individual Differences in Attention Control and Allocation of Attention Across the Visual Field

A Dissertation
Presented to
The Academic Faculty

by

Christopher Draheim

In Partial Fulfillment of the
Requirements for the Degree
Doctor of Philosophy in the
School of Psychology

Georgia Institute of Technology
December 2021

Change Detection as a Framework for Understanding Individual Differences in Attention Control and Allocation of Attention Across the Visual Field

Approved by:

Dr. Randall W. Engle, Advisor
School of Psychology
Georgia Institute of Technology

Dr. Paul Verhaeghen
School of Psychology
Georgia Institute of Technology

Dr. Rick P. Thomas
School of Psychology
Georgia Institute of Technology

Dr. Michael J. Beran
Department of Psychology
Georgia State University

Dr. Christopher Hertzog
School of Psychology
Georgia Institute of Technology

Date Approved: October 18, 2021

ACKNOWLEDGEMENTS

I thank all lab members, undergraduate, graduate, and post-doctoral, for their assistance in data collection, data aggregation, and overall study management. In particular, I would like to thank Mariel Barnett for her tireless work on the study at large. I also thank the crucial efforts of Alex Burgoyne, Cody Mashburn, and Jason Tsukahara.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	**iii**
LIST OF TABLES	**vi**
LIST OF FIGURES	**vii**
LIST OF ABBREVIATIONS	**viii**
SUMMARY	**ix**
CHAPTER 1. Introduction	**1**
1.1 Theory of Individual Differences in Attention Control	**1**
1.1.1 Is Attention Control Multifaceted?	5
1.2 Measurement of Attention Control	**7**
1.2.1 Challenges in Assessing Individual Differences in Attention Control	9
1.3 Allocation of Attention Across the Visual Field	**24**
1.3.1 Types of Allocation and Individual Differences	24
1.3.2 Limitations and Gaps in the Literature	29
CHAPTER 2. The Present Study	**32**
2.1 Method	**33**
2.1.1 Participants and General Procedure	34
2.1.2 Tasks of Interest	35
2.1.3 Data Preparation	56
2.2 Results	**58**
2.2.1 Task Abbreviations	58
2.2.2 Piloting	58
2.2.3 Reliability	60
2.2.4 Descriptive Statistics	60
2.2.5 Intercorrelation and Factor Structure	62
2.2.6 Predictive Validity and Relationship Among Factors	69
2.2.7 Differences in Performance Across Set Sizes	77
2.2.8 Effect of Distractors	82
2.2.9 Feature vs. Location Distraction	84
2.2.10 Differences in Spatial Configuration of Stimuli	89
2.2.11 Differences in Performance Across Preparation Time	91
CHAPTER 3. General Discussion	**99**
CHAPTER 4. Conclusion	**100**
APPENDIX A. ADDITIONAL INFORMATION	**101**

A.1 Table A1 **101**
A.2 Table A2 **101**

REFERENCES **102**

LIST OF TABLES

Table 1	Descriptive Statistics for the Visual Arrays Tasks	61
Table 2	Zero-Order Correlations Among Visual Arrays Tasks	62
Table 3	Semipartial Correlations between Attention Control and Visual Arrays Regressed on Various Criteria	75
Table 4	Semipartial Correlations between Processing Speed and Visual Arrays with Various Criteria	77
Table 5	Comparing Set Size 3 and Set Size 5 Performance to Various Criteria	80
Table 6	Incremental Criterion Validity by Adding Distractors	83
Table 7	Semipartial Correlations between VA4 and CVA on Several Criterion Measures	86
Table 8	Semipartial Correlations between EVA4 and ECVA on Several Criterion Measures	87
Table 9	Semipartial Correlations for Factor Scores of Color- and Location-Based Selection	87
Table 10	Semipartial Correlations for Spotlight and Donut Trials	91
Table 11	K Scores as a Function of Cue-to-Stimulus Interval	94
Table 12	Semipartial Correlations of Different Cue-to-Stimulus Intervals	95
Table A1	Full List of Session 1-4 Tasks in the Study at Large	101
Table A2	Zero-Order Correlations Among Visual Arrays Tasks and Other Constructs	101

LIST OF FIGURES

Figure 1	The Maintenance and Disengagement Framework	3
Figure 2	Attention Control as a Full Mediator of the WMC-Gf Relationship	5
Figure 3	Attention Control and Processing Speed Predicting WMC and Gf	19
Figure 4	The Selective Visual Arrays Task from Draheim et al. (2021b) and Shipstead et al. (2014)	20
Figure 5	The Complex Span Tasks	37
Figure 6	Trial Sequence for the Sustained Attention-to-Cue Task	42
Figure 7	Non-Selective Visual Arrays (NSVA)	45
Figure 8	Color Selective Visual Arrays (VA4)	46
Figure 9	Enhanced Color Selective Visual Arrays	48
Figure 10	Concentric Selective Visual Arrays (CVA)	51
Figure 11	Practice Figures for the Concentric Selective Visual Arrays Task	52
Figure 12	Enhanced Concentric Visual Arrays Task (ECVA)	53
Figure 13	Practice Figures for the Enhanced Concentric Selective Visual Arrays Task	54
Figure 14	Histograms of Visual Arrays Performance	61
Figure 15	Factor Structure with Each Visual Arrays Task Entered Separately	67
Figure 16	Factor Structure with Visual Arrays Entered Together	68
Figure 17	Confirmatory Factor Analysis Showing Relationship Among Factors	70
Figure 18	Attention Control Mediating the WMC-Gf Relationship	72
Figure 19	Attention Control and Visual Arrays Predicting WMC and Gf	73

LIST OF ABBREVIATIONS

WMC	Working Memory Capacity
Gf	Fluid Intelligence
AC	Attention Control
NSVA	Non-Selective Visual Arrays
VA4 or VA	Color Selective Visual Arrays
EVA4 or EVA	Enhanced Color Selective Visual Arrays
CVA	Concentric Visual Arrays
ECVA	Enhanced Concentric Visual Arrays
VA-S	Selective Visual Arrays
SACT	Sustained Attention-to-Cue
RAPM	Raven’s Advanced Progressive Matrices
RotSpan	Rotation Span
SymSpan	Symmetry Span

SUMMARY

Attention control is a domain-general ability that guides the control of thoughts and intentional behavior in a goal-driven manner and is a central concept to many models of human cognition. Our lab recently showed that attention control can be measured more reliably and validly with alternative tasks, one of which being selective visual arrays (rapid change detection with distracting stimuli). The present study was designed with two goals in mind: first, to extend this finding by further exploring the nature of attentional individual differences in visual arrays tasks, and second to use the visual arrays paradigm to investigate individual differences in how individuals allocate attention across the visual field. Five variants of visual arrays were administered to 210 participants from the Atlanta community along with a battery of other cognitive tasks. Results showed that the presence of distractors in visual arrays was the most important factor in scores producing attention related individual differences. Further, variants that had more distractors and/or required spatial selection of targets, as opposed to feature selection, were more difficult and more strongly predictive of overall cognitive ability. On the other hand, (1) performance on supra-capacity vs. near-capacity array sizes were not differentially predictive of cognitive ability, (2) within the variants that required spatial selection of targets there were no substantive differences in performance as a function how the targets and distractors were arranged, and (3) there were no detectable meaningful differences in performance across different cue-to-stimulus intervals. In the discussion section I explore how and potentially why some of these results are consistent with my hypotheses whereas some were unexpected and thus contrast with findings from the literature. The overall conclusion is

that the visual arrays paradigm is an attention control measure robust to a variety of manipulations.

CHAPTER 1. INTRODUCTION

Executive attention plays a central role in most models of higher-order cognition (Atkinson & Shiffrin, 1968; Baddeley & Hitch, 1974; Botvinick et al., 2004; Egeth & Yantis 1997; Norman & Shallice, 1986; Posner & DiGirolamo, 1998; Shipstead et al., 2016). Broadly defined, executive attention guides the control of thoughts and behavior with intention in a goal-driven manner and is particularly important when there is conflict between more automatic processes and one's intentions. It has been shown that individual differences in executive attention, which will henceforth be referred to as *attention control*, predict higher-order cognitive abilities (Engle, 2002). Further, attention control is important for many every-day behaviors and real-world phenomena, including self-control (Broadway et al., 2010), emotional regulation (Schmeichel & Demaree, 2010), and task engagement (Miller & Cohen, 2001; Botvinick et al., 2004; see also Draheim et al., 2021a).

1.1 Theory of Individual Differences in Attention Control

Engle and colleagues have argued that ability to control attention accounts for much of the individual variation in working memory capacity (WMC) and fluid intelligence (Gf) task performance (e.g., Kane et al., 2007; Mashburn et al., 2020; Shipstead et al., 2012) and that attention control is a domain-general ability important for cognitive functioning (Kane et al., 2001). Still, the mechanisms involved in attention control remains an open question. Here, I will operate under the framework that attention control is a broad, domain-general, ability to maintain goal-directed behavior, particularly in the face of distraction and/or cognitive interference (see Engle & Kane, 2004). This is very similar to our lab's conception of WMC, however the critical difference is that WMC specifically involves maintaining and manipulating goal-relevant *information* in primary memory (which draws resources from a top-down executive system), whereas attention control refers to how limited-capacity attention is used to manage goal-directed behavior more broadly, which may or may not specifically involve the maintenance of information.

To further elaborate on the distinction between WMC and attention control, Shipstead et al. (2016) proposed that the strong relationship between WMC and Gf (roughly 50% shared variance; see Ackerman et al., 2005; Engle et al., 1999; Kane et al., 2005; Oberauer et al., 2005) is largely due to the tasks of each construct requiring two distinct, but complimentary, mechanisms – maintenance of information in the face of distraction, and disengaging from previously relevant but now irrelevant information. In this framework, maintenance and disengagement both rely on a top-down executive attention system and therefore operate in tandem to facilitate goal-directed behavior. This top-down executive attention system is therefore the common resource for performing WMC and Gf tasks and is how attention control is implemented. That is, attention control regulates both the maintenance of information (and behavior) and the filtering, blocking, and disengaging from irrelevant information (and inappropriate behavior). While the reliance on attention control in the form of maintenance and disengagement explains the strong relationship between WMC and Gf, the two constructs are not isomorphic. This is hypothesized to be because a) WMC and Gf tasks involve more than maintenance and disengagement, and b) WMC and Gf involve maintenance and disengagement to differing degrees. Specifically, WMC tasks are thought to place more demand on the maintenance of information in primary memory whereas Gf tasks place more demand on disengaging from irrelevant information, for example resisting perseveration and not retesting the same failed or outdated hypothesis, e.g., while performing a matrix reasoning problem (Figure 1; also see Martin et al., 2020). This may explain, for example, why working memory "updating" tasks (e.g., running span and mental counters) correlate more strongly with Gf measures than do complex span tasks (e.g., Colom & Shih, 2004; Shipstead et al., 2014); relative to complex span, updating the contents of primary memory requires more disengaging from previously relevant but now irrelevant information.

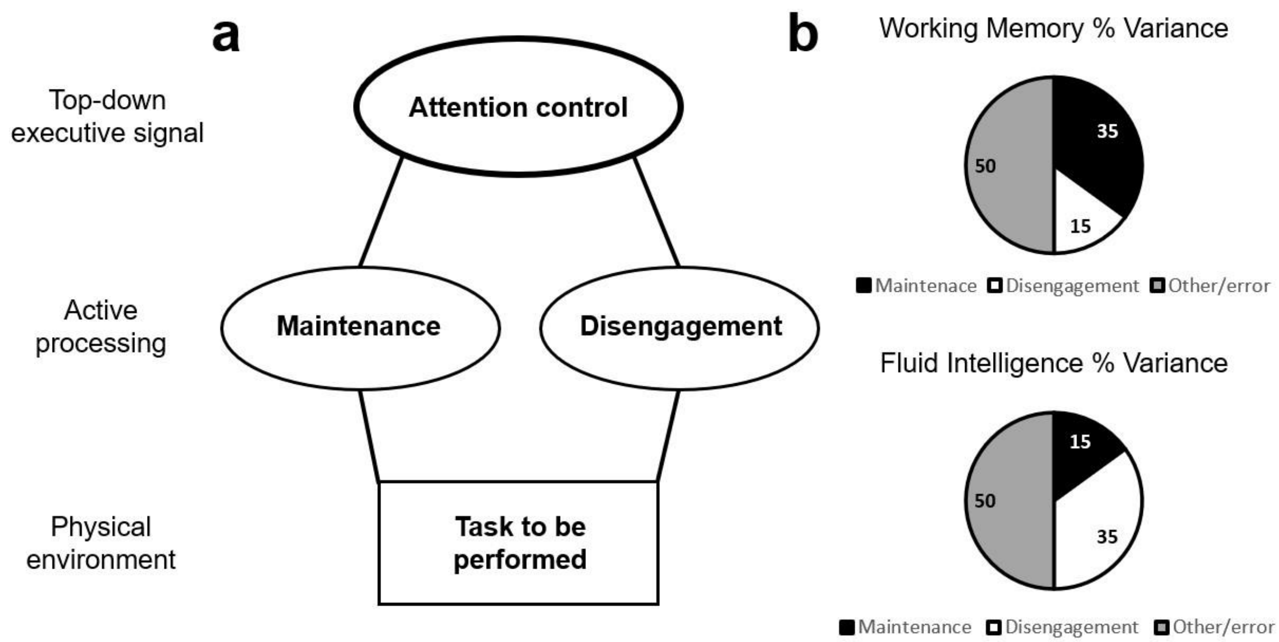

Figure 1 – The Maintenance and Disengagement Framework. *Note.* (a) Illustration of the maintenance/disengagement hypothesis proposed by Shipstead et al. (2016). Maintenance and disengagement are hypothesized to be separate but complimentary processes that engage an overarching and top-down executive attention system (attention control) in performing cognitive tasks. (b) A hypothetical breakdown of performance variance in WMC and Gf tasks showing the relative difference in requirement for maintenance of information as opposed to disengagement from information.

In this theoretical position, attention control is the top of the hierarchy for cognitive functioning and should account for a substantial amount of individual variation in most, if not all, higher-order cognitive tasks and behaviors (see Burgoyne & Engle, 2020; Draheim et al., 2021a). This is compatible with Kovacs and Conway's (2016) process overlap theory account of intelligence, of which "A central claim … is that domain-general executive attention processes play a critical role in intelligence, acting as a central bottleneck on task performance and a constraint on development of domain-specific cognitive abilities" (p. 2; Conway et al., 2021). This view is also consistent with Rueda's (2018) argument that the regulation of attention is at the heart of intelligence and therefore underlies higher-order cognitive functioning. This position was motivated in part by evidence showing that higher-order cognitive tasks activate the same frontal network that is associated with attention control, and that the expansion of the anterior cingulate cortex (identified as an important area for attentional behavior) appears to have been critical in the evolution of human intelligence. Recent factor analytic studies from our lab have sought, and uncovered, evidence for these claims. For example, Draheim et al. (2021b) found that the strong relationship between WMC and Gf was no longer statistically significant once an attention control factor was added to the model as a mediator (Figure 2). Likewise, Tsukahara et al. (2020) found that attention control fully mediated the nearly 50% shared variance between factors of WMC and sensory discrimination ability.

If we assume that the field's conception of inhibition is functionally subsumed by our conception of attention control, then this is in stark contrast to the various views recently expressed by researchers that inhibition is task-specific, not a unified construct, or that there is “nothing special” about it (e.g., Friedman and Miyake, 2017; Rey-Mermet et al., 2018). The reasons for these dissenting views will be discussed in following sections.

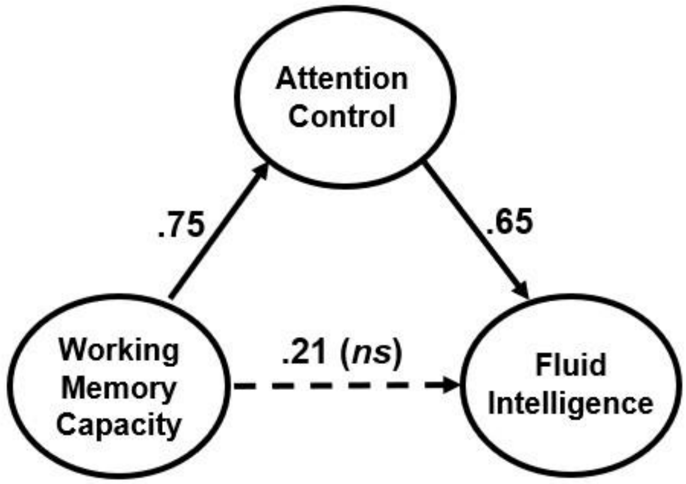

Figure 2 – Attention Control as a Full Mediator of the WMC-Gf relationship. *Note*. Structural equation model from Draheim et al. (2021b) showing that the relationship between working memory capacity and fluid intelligence is not statistically significant when attention control is added as a mediator. Each construct was measured with three tasks. N = 396.

1.1.1 Is Attention Control Multifaceted?

An open question is whether attention is comprised of a subset of separable abilities or mechanisms (multifaceted), or if instead that attention control is a unitary construct. There are a wide number of perspectives regarding the nature and number of attentional facets. Just to give some examples, Posner and colleagues have argued for the existence of three independent attentional “networks” assessed by the popular Attention Network Test: alerting (preparing for a stimulus by establishing and maintaining alertness), orienting (shifting attention to an incoming stimulus), and executive attention (detection and resolution of cognitive conflict; Fan et al., 2002 Posner & Peterson, 1990). Kane et al.

(2016) distinguished between tasks that require attentional restraint (such as resisting a prepotent response in the Stroop or antisaccade) vs. attentional constraint (such as narrowing the focus of visual attention to a small space in a flanker task). Friedman and Miyake (2004) hypothesized three types of inhibitory functions: 1) inhibiting a prepotent response, 2) resisting distraction interference, and 3) resisting proactive interference. And, recently, Unsworth and colleagues have found value in distinguishing between the intensity (mental effort) vs. the consistency (stability) of attention, supported in part by individual differences in the magnitude of pupil dilation (intensity) and fluctuation of pupillary response (consistency) while performing demanding cognitive tasks (e.g., Unsworth & Miller, 2021).

In an attempt to reconcile these taxonomies with our lab's conception of attention control as domain-general and unitary ability, the approach I take here is to treat attention control as a unitary psychometric construct but with the caveat that it can be applied differently depending on the task at hand. In other words, one task or situation may place demands on the individual to control attention in a different manner than another, but this does not necessarily mean that these situations reflect different mechanisms. To provide an example, the psychomotor vigilance task requires sustaining attention over a period of time whereas the antisaccade requires resisting a strong, evolutionary engrained, prepotent response to orient to movement in the periphery, but these tasks require attention to be applied differently rather than reflecting two distinct attentional mechanisms. In support of this view, in Draheim et al. (2021b) we found that a collection of ten existing, modified, and new measures hypothesized to tap attention control generally loaded onto a single latent factor that predicted substantial variance in WMC and Gf above and beyond other cognitive tasks and latent variables (e.g., processing speed), but also that in some analyses a subset of the attention measures (specifically, flanker tasks) clustered together and separated from the single factor, as indicated by strong loadings on a secondary factor.

1.2 Measurement of Attention Control

Given that attention control plays such a prominent role in models of higher-order cognition, one would assume that the investigators have an array of reliable and valid methods to measure it. In experimental psychology, this is perhaps true. Established paradigms such as Stroop and flanker come in numerous forms and are remarkably robust, with the Stroop effect in particular being considered universal and perhaps the most studied and well-replicated phenomenon in experimental psychology (MacLeod, 1991; Verhaeghen & De Meersman, 1998). However, from the differential perspective, this is unequivocally not the case. For instance, the robust and universal Stroop effect has demonstrably poor reliability and predictive validity and tasks designed to measure attention control (often conceptualized as inhibitory control), generally share under 4% of their variance at the task level (Draheim et al., 2019; Rouder & Haaf, 2019; Rouder et al., 2019). Compare this with as high as 25-50% for tasks of established cognitive constructs such as working memory capacity and fluid intelligence (e.g., Kane et al., 2001).

Problems with the assessment of individual differences in attention control were shown quite clearly in 2004. Using a sample of 220 university students, Friedman and Miyake (2004) attempted to distinguish three different types of attention control each with three indicators: resisting a prepotent response, resistance to distractor interference, and resistance to proactive interference. At the task level, reliabilities were below .80 for seven of their nine primary attention measures, most intercorrelations were statistically nonsignificant, and confirmatory factor analysis loadings ranged from .29 to .55 with an average below .40. Friedman and Miyake (2004) concluded:

> …for most so-called inhibition tasks, the relative proportion of the variance attributable to the hypothesized inhibition ability may be quite small in comparison with the variance attributable to other idiosyncratic requirements of the task or the error variance…One obvious solution to this problem is to develop new tasks that are more psychometrically reliable and more sensitive to individual variation in inhibition-related processes…it is becoming increasingly clear that new measures are needed for the field to make further progress. Such measures must be relatively simple and easy to administer, demonstrate high reliability, and primarily tap one of the inhibition-related functions examined here or hypothesized in the literature. More important, such tasks must be able to tap more inhibition-related variance than has been possible with the existing measures. (p. 127)

Despite this call-to-arms, many of the same problematic measures of attention have seen continued use in correlational studies and usually with results not much better than Friedman and Miyake (2004; e.g., Draheim et al., 2019; Friedman & Miyake, 2017; Hedge et al., 2018; Rey-Mermet et al., 2018; Rouder & Haaf, 2019; Rouder et al., 2019). Friedman and Miyake (2017) noted that when inconsistencies arise in studies from other researchers that adopt their 3-executive function framework (updating, inhibition, and shifting; Miyake et al., 2000), the problems are usually with the inhibition factor, that is the attention control measures. Specifically, inhibition does not separate from the other factors and/or correlations between inhibition tasks are too weak to produce a coherent latent variable. It is no wonder then why some researchers argue that attention control is a task-specific mechanism that has relatively little value or utility for explaining human cognition in a broad sense (e.g., Rey-Mermet et al., 2018; Whitehead et al., 2020).

However, there has recently been a push for sweeping changes from several independent groups of researchers (e.g., Draheim et al., 2019; 2021b; Hedge et al., 2018; in press; Paap & Sawi, 2016; Rey-Mermet et al., 2019; Rouder et al., 2019). A growing number of researchers are starting to recognize the task-impurity and reliability issues of attention measures pointed out by Friedman and Miyake, with the largest offenders being the aforementioned conflict-resolution tasks (Stroop, flanker, and Simon). And while researchers do not all agree on the cause of the problem or best solution, it is becoming increasingly recognized that theoretical advances have stagnated and *something* needs to be done to reinvigorate behavioral research of individual differences in attention.

1.2.1 Challenges in Assessing Individual Differences in Attention Control

There are several non-mutually exclusive explanations and perspectives for the consistent finding that individual differences in attention control tasks are generally uncorrelated. I will only briefly cover some of these perspectives, but see von Bastian et al. (2020) for a more thorough review of the state of affairs within the assessment of individual differences in attention control.

It warrants mention that there is a critically important yet often overlooked difference between experimental (group-level/aggregate) approaches to research vs.

differentially (correlational/individual differences) approaches to research. Tasks such as the Stroop task continue to experience widespread use among attention researchers because of how established the Stroop effect is and the large body of experimental research behind it (see MacLeod, 1991). But reliability for experimental and differential researchers is not the same (see Cronbach, 1957), and therefore tasks suited to experimental studies are not necessarily suited to correlational pursuits (see Draheim et al., 2019; Goodhew & Edwards, 2019; Hedge et al., 2018; Logie et al., 1996). Hedge et al. (2018) astutely argued that it might be the very characteristics that make tasks ideal for experimental research (such as minimization of between-subject variance) that are responsible for them being poor for correlational purposes, explaining why the Stroop task in particular is so problematic. Hedge et al. dubbed the trend that robust cognitive tasks do not produce reliable individual differences the “reliability paradox.”

Proposed reasons that attention tasks do not cohere well include: (1) task unreliability, (2) task impurity, and (3) that attention control is not a unitary concept or ability.

In terms of reliability, attention measures are often scored through the use of difference scores, which are known to be necessarily less reliable than their component scores in practical settings (Cronbach & Furby, 1970; Lord, 1956, also see Draheim, et al., 2019 and Hedge et al., 2018). As a result, reliability of these interference effects derived from these tasks is highly attenuated, with various estimates that anywhere from 34 – 60% of the variance in these tasks is strictly error variance (e.g., von Bastian et al., 2020;

Draheim et al., 2019; Hedge et al., 2018; Friedman & Miyake, 2004; Paap & Sawi, 2016; Rey-Mermet et al., 2018).[1] To wit, attention control tasks generally have relatively low effect sizes and little between-participants variance relative to trial-level noise, thus resulting in low reliability and validity. The following paragraphs will discuss some issues with difference scores in the context of reaction time (RT), as RT differences are the most common way to score many cognitive measures, but it is important to note that accuracy-based difference scores are also problematic and generally for the same reasons (see Hughes et al., 2014; Draheim et al., 2019).

Beyond reliability concerns, several sources of contamination have been proposed for attention measures, with many researchers focusing on the ubiquitous Stroop, flanker, and Simon tasks. For example, Verhaeghen and De Meersman (1998) and Rey-Mermet et al. (2019) noted that RT difference scores are disproportionally inflated by individual differences in speeded responding or processing speed. That is, individuals who are faster will have smaller interference effects in tasks such as Stroop and flanker simply because they are faster overall, and not because they are better at resolving cognitive interference or conflict. Similarly, in a recent study by Hedge et al. (in press) the authors argued that performance in the Stroop, flanker, and Simon tasks is contaminated with variance attributable to strategy (specifically, speed-accuracy emphasis) and processing speed. They analyzed seven data sets using the diffusion model for conflict tasks (Ulrich et al., 2015)

[1] Reported reliability indices may also be inflated as it is sometimes unclear how reliabilities are calculated and, in many cases, they appear to be calculated incorrectly such that reliability is calculated in a way that does not match the outcome scores. For instance, some researchers report the average reliability of the component scores (e.g., average reliability of mean RT on incongruent trials and mean RT on congruent trials in a Stroop task) instead of calculating reliability specifically for the difference between these conditions. The field-wide problem with calculating reliability incorrectly has been noted by some researchers (e.g., Parsons et al., 2019) and is a threat to properly assessing reliability of these measures.

and found that model parameters associated with conflict resolution (attention control) did not correlate across tasks, whereas drift rate and boundary parameter parameters associated with construct-irrelevant processes (general processing speed/efficiency and response cautiousness, respectively) were strongly correlated across tasks. In a simulation, they also showed that even if they imposed an artificial correlation among conflict parameters between pairs of tasks, it would generally not manifest as strong correlations among RT or error interference effects between the tasks. Whereas a strong correlation between non-conflict parameters would result in relatively strong correlations in these interference effects across tasks. They concluded that (1) conflict tasks reflect virtually no shared variance associated with the mechanisms they are believed to measure; (2) conflict tasks share substantial variance associated with processing speed and speed-accuracy emphasis; and (3) scores on these tasks are minimally informative regarding the debate as to whether inhibition/attention control is task-specific or a unitary concept. In other words, due to the tasks being invalid, weak correlations do not provide evidence against attention control being unitary, whereas hypothetically strong correlations would not provide evidence that attention control is a unified ability. Keep this in mind for the following paragraph.

A study that has gained significant attention among researchers is Rey-Mermet et al. (2018) in which the authors employed eleven commonly used attention tasks (including five Stroop, flanker, and Simon tasks) in a diverse sample of 289 young and old participants and found little shared variance among the tasks. Specifically, only 25% of their task-level correlations were statistically significant; only 11% of their correlations exceeded $r = .20$; and when the tasks were loaded onto a common factor at the latent level only, five had factor loadings above .18 (far from acceptable even if one adopts a liberal tolerance; e.g., Comrey & Lee; 1992; Matsunaga, 2010; Stevens, 1992). Of note is that all but one of their measures (antisaccade) were scored as an RT difference, and the average internal consistency of the RT difference score tasks was only .50. The reason this study is noteworthy is not necessarily due to the unsurprising results, but because of their recommendation that researchers should stop thinking about attention control as a general cognitive construct. Instead, they argued that their results strongly supported the task specific nature of the conflict resolution mechanisms involved in these tasks. The following year, Rey-Mermet et al. (2019) noted that RT difference scores can be artificially inflated due to differences in processing speed/efficiency (see Verhaeghen & De Meersman, 1998), which was one of the concerns expressed by Hedge et al. (in press). They therefore tested whether accuracy-based measures of attention would cohere together and correlate to WMC and Gf. Their results were no better than Rey-Mermet et al. (2018), however of note is that they still relied on difference scores and that they employed a novel calibration procedure (see Draheim et al., 2021b).

Miyake, Friedman, and colleagues have also argued that attention control is not a separable and unified ability. In several datasets (from their lab and others) they noted that attention measures have low task-level correlations, fail to separate from latent variables of other executive functions, and/or do not produce a coherent latent factor (see Friedman & Miyake, 2017). Specifically, they showed that in studies assessing multiple executive functions, attention control factors are difficult to extract largely because a common executive functioning latent is sufficient to account for relationships involving inhibition tasks (e.g., Miyake & Friedman, 2012). And although some take this as evidence that inhibition may be the common underlying mechanism (or set of mechanisms) for all executive functioning, they interpret these results as evidence that there is nothing special or unique about inhibition.

1.2.1.1 What is the Problem with Attention Tasks?

Several recent papers from our lab have outlined our position on the issue of why many cognitive tasks are not optimal for studying individual differences and potential solutions to this issue (e.g., Draheim et al., 2016; 2019; 2021b; Martin et al., in press). To summarize our view: (1) difference scores are poorly suited to correlational work because they have lower reliability and do not isolate processes of interest as widely believed; (2) tasks that place demands on *both* accurate and quick responding are susceptible to contamination from construct-irrelevant variance (namely processing speed and speed-accuracy interactions); and, (3) new paradigms and/or modifications of traditional paradigms are necessary to advance the field.

In my view, differences scores are problematic not only because they are less reliable but because they are assessed such that only RT is considered (see Draheim et al., 2016; 2020). As to why this is a problem, first consider Hedge et al. (in press) in which the authors argued that there are multiple sources of variance captured by difference scores in RT in Stroop, flanker, and Simon tasks. The first source is due to ability, such that an individual with better attention control should (ceteris paribus) have smaller RT costs and smaller error costs. The second source is due to quicker processing speed which would result in lower RT costs and (arguably) lower error costs as well. But the third source, response cautiousness, will result in *higher* RT costs but *lower* error costs. That is, an individual who emphasizes accuracy and sacrifices speed (which may be due to their baseline tendency, in response to task instructions or demands, or as a trial-by-trial adjustment due to feedback or their own performance monitoring) will slow down more on the attention-demanding trials (e.g., incongruent Stroop trial) to maintain a similar level of accuracy as on the baseline trials (e.g., congruent Stroop trial). This speed-accuracy tradeoff was present in a dataset from our lab in RT costs from a task switching paradigm, and manifested as a positive correlation between RT switch costs and WMC/Gf (Draheim et al., 2016). That is, those who had larger RT switch costs (which ostensibly reflects poorer task-switching ability) also performed better on the WMC and Gf measures.

The larger issue here is that if a task requires respondents to respond both quickly and accurately, using RT as the dependent variable can be problematic independent of whether difference scores are used (Draheim et al., 2019). This is supported by research showing that individual differences in speed-accuracy tendencies are shared across cognitive tasks, including studies analyzing attention measures (e.g., Hedge et al., 2019; in press; Starns & Ratcliff, 2010; Whitehead et al., 2020). Several researchers have argued that *both* speed and accuracy need to be considered for scores to be a valid reflection of what they are intended to measure. For example, Wickelgren (1977) argued that in almost all cases, researchers should use speed-accuracy methodology over RTs, and Luce (1986) argued that the only sensible way to combat the speed-accuracy tradeoff problem was to study it and devise a summary statistic to measure it. And while speed-accuracy differences along the developmental continuum are well documented and often considered in aging research (e.g., Forstmann et al., 2011; Hertzog et al., 1993; Starns & Ratcliff, 2010), researchers studying healthy young adults tend to be less mindful of speed-accuracy interactions, which can manifest in a number of different ways and interact with ability (for a review, see Heitz, 2014).

1.2.1.2 The Toolbox Approach to Measuring Attention Control

In a large-scale study using a latent variable approach, Draheim et al., (2021b) took the approach to administer attention tasks that were designed to control either for RT or for accuracy. We administered modified Stroop and flanker tasks which were adaptive based on either a response deadline or presentation rate, developed new accuracy-based and adaptive tasks, and included several tasks in which RT or accuracy was the dependent variable but the other was irrelevant. The antisaccade task is one example and is the only attention control measure that has consistently proven to be reliable, have strong factor loadings to an attention latent variable, and correlates strongly to other cognitive measures in our studies (as opposed to RT costs in Stroop and flanker), and our previous correlational studies of attention control relied heavily on the strength of this task (e.g., Redick et al., 2016; Shipstead et al., 2014; 2015). The same can be said for other researchers as well, as Rey-Mermet et al. (2019) observed across a number of independent studies that antisaccade tasks (see the Tasks of interest section for a description) tend to "dominate" latent factors of attention such that factor loadings are high with this task and low, often unacceptably so, with other attention-related measures when researchers attempt to create a latent variable of attention. The reasons we believe this task is such a good individual differences measure are because it is simple, not scored using contrasts[2], and places no demands on

[2] Note that there are different versions of antisaccade, some with involve RTs and/or taking a difference score between performance on baseline (prosaccade) trials and antisaccade trials. Not all labs have success with this task, and it could be due to differences how the task is administered or scored (c.f., Draheim et al., 2021a).

responding quickly. In other words, this task avoids many of the issues that other attention tasks have.

The adaptive and accuracy-based measures were overall more reliable and intercorrelated much more strongly than the standard RT cost versions of the Stroop and flanker. Relative to the standard RT difference-score versions of the Stroop and flanker that were included in the study for comparison, the accuracy-based tasks had 5x as much reliable and valid between-participants variance and the adaptive Stroop and flanker tasks had roughly 3x as much. But the highlight was that we found strong coherence among almost all the accuracy-based measures, some with task-level correlations as strong as r = .45 despite the tasks sharing few superficial similarities. These tasks also cohered at the latent level much better than the typical attention measure, and we even found that these accuracy-based attention tasks could *fully* mediate the relationship between WMC and Gf at the latent level (refer to Figure 2). This finding is quite remarkable because, even though our lab's theory that attention control is the link between WMC and Gf would predict that attention control fully explains the WMC-Gf relationship, this finding had been elusive up to that point. For example, some researchers have argued that secondary memory, memory storage, *and* attention control each uniquely contribute to the WMC-Gf relationship and so any one ability cannot fully account for the relationship on its own (e.g., Shipstead et al., 2014, Unsworth et al., 2014; Unsworth & Spillers, 2010). And in the dataset from Draheim et al. (2021b), standard RT-based attention measures (Stroop, flanker, and psychomotor vigilance) accounted for less than half of the total shared variance between WMC and Gf at the latent level. On the other hand, the accuracy-based attention measures failed to account for only 3-4% of the variance in the WMC-Gf relationship, which was not statistically different from 0% (full mediation).

Because we avoided using contrasts (difference scores) in developing and scoring the attention control tasks in Draheim et al. (2021b), it could be argued that our so-called attention tasks were not process pure and thus reflected a large amount of variance not attributable to attention control. In other words, increased coherence and predictive validity is theoretically meaningless if it is due to, say, contamination from processing speed. Importantly, however, we included processing speed tasks in this study, which allowed us to directly test whether these attention measures were contaminated with processing speed. We found that although the processing speed measures correlated with the attention measures (sharing 40% of their variance at the latent level), at the latent level processing speed contributed no unique variance to WMC and Gf above and beyond the strongest performing attention tasks, whereas attention control predicted *substantial* incremental variance in WMC and Gf above and beyond processing speed (38% to WMC and 41% to Gf; see Figure 3).

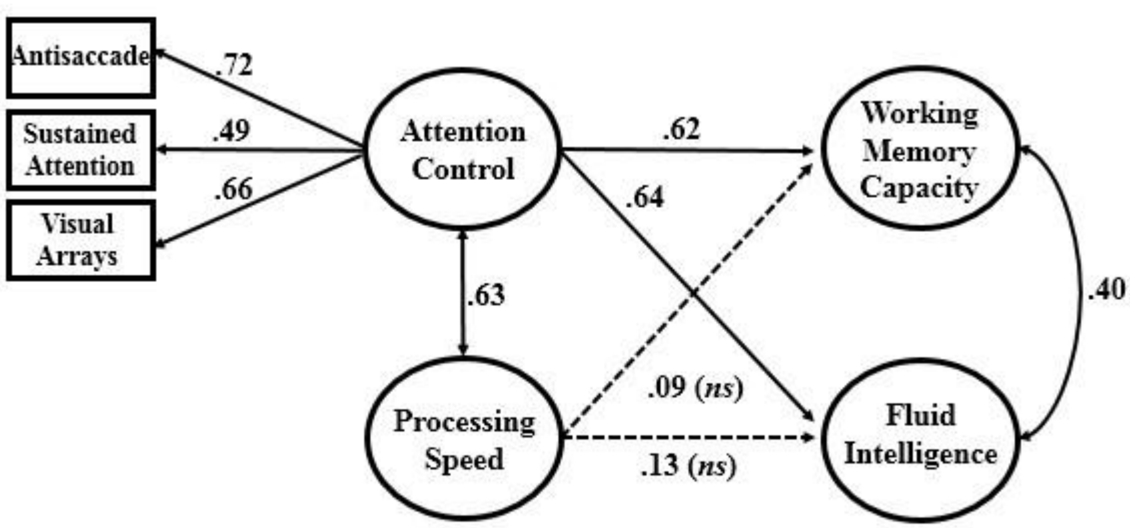

Figure 3 – Attention control and processing speed predicting WMC and Gf. Structural equation modeling with correlated attention control and processing speed predicting WMC and Gf. Data from Draheim et al. (2020). Attention control is comprised of the three best performing tasks (out of ten) according to the assessed criteria, which happened to be the three non-adaptive accuracy-based measures. Dashed lines indicate non-statistically significant paths. *N* = 173.

1.2.1.3 <u>Selective Visual Arrays as an Indicator of Attention Control</u>

In Draheim et al. (2020b), we used four criteria for performance to assess our measures of attention control. They were (1) test-retest reliability, (2) intercorrelation to other attention measures, (3) factor coherent on an attention latent, and (4) relationship to WMC and Gf. We found that the version of the change detection task we administered (selective visual arrays; see Figure 4 for a description) was on average the second strongest measure (just behind antisaccade) across our four criteria in terms of the amount of reliable and shared variance. However, the categorization of visual arrays as an attention measure was contentious at the time. Even though the ability to detect change in the environment is widely considered an attentional phenomenon (e.g., Rensink et al., 1997; for neuropsychological evidence, see Huettal et al., 2001), the visual arrays paradigm is largely considered to be a measure of visual working memory capacity because it requires holding multiple memoranda in primary memory. So, the critical view of the results from Draheim et al. would be that we found a strong relationship between attention control and WMC because we simply took a WMC measure and erroneously categorized it as attention control task.

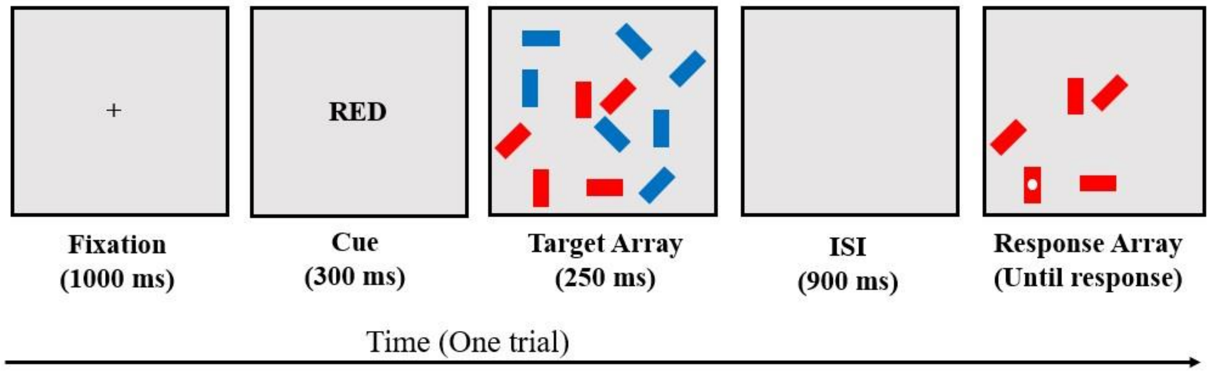

Figure 4 – ***The Selective Visual Arrays Task from Draheim et al. (2021b) and Shipstead. et al. (2014)*** *Note.* On each trial, the participant is briefly cued to which color of rectangles to attend to (either red or blue) prior to a stimulus array of 12 rectangles, with either 5 or 7 being distractors, for 250 ms. After a 900 ms delay, the stimulus display appears with only the rectangles of the cued color. One of which has a white mark on it, and the participant's goal is to judge whether this rectangle has changed orientation from the initial array. The rectangles can be horizontal, vertical, diagonal at 45 degrees, or diagonal at 145 degrees. A total of 80 trials were administered, 40 of each set size. The dependent variable is the average capacity (k) score, calculated using the method outlined by Cowan et al. (2005). This task is considered *selective* because of the additional filtering demand placed on the respondent to selectively attend to rectangles of the cued color and ignore rectangles of the non-cued color. Shown is set-size 5 as there are five target (cued) rectangles. Schematic is not to scale, as the rectangles were enlarged for clarity in this figure and occupy much less space on the screen as shown here.

The rationale for categorizing visual arrays as an attention measure is laid out in Martin et al. (in press). It is first important to distinguish between two types of visual arrays tasks: non-selective and selective. As the name implies, the critical difference is whether there is an additional requirement for the respondent to select a subset of stimuli to retain (selective) or whether the respondent instead should encode and be prepared to recall all stimuli in the original display (non-selective). Selection in this case requires attention to be focused on one source or kind of information at the exclusion of others and therefore may be accomplished through enhancing attention on the target stimuli, filtering out or disengaging from distracting stimuli, or a combination of these processes (e.g., Davies et al., 1984). This selection, or filtering, component is implemented as a brief cue prior to the first array presentation that indicates which stimuli are to be maintained. Individuals who do not properly utilize this cue will effectively try to attend to more items than individuals who perfectly use this cue and select the appropriate stimuli, resulting in worse performance for those who do not use attention control to focus on the cued material and block, inhibit, ignore, or disengage from the distracting stimuli. Behavioral and electroencephalogram results show that capacity estimates (*k*) for selective tasks are generally much lower than the *k* = 4 or so typically observed in the non-selective visual arrays tasks (Fukuda et al., 2015; Shipstead et al., 2014). This does not *necessarily* indicate that individuals retain less total information in the selective versions, but instead some evidence shows that roughly the same total number of stimuli are encoded and retained in both versions but that a percentage of the retained stimuli in the selective versions are the distractors, resulting in a lower *k* score (e.g., Fukuda & Vogel, 2009).

There is evidence that non-selective visual arrays tasks also require a good deal of attentional resources (e.g., Shipstead et al., 2014). For example, Cowan and Morey (2006) found that a selective attention measure explained a large proportion of the shared variance between non-selective visual arrays and intelligence. Fukuda and Vogel (2011) found that low WMC individuals took longer to recover from attentional capture in non-selective visual arrays, and that this was a source of individual variation of overall task performance. And Fukuda et al. (2015) reported that *k* scores in visual arrays tasks fluctuate very little for high-WMC individuals as a function of set size (e.g., *k* scores are the same for set-size 4 and set-size 7) whereas low-WMC individuals have increasingly smaller *k* scores as set sizes increase. They reasoned that when set sizes exceed capacity limits, attention control is required to properly allocate WM resources, and thus individuals with poorer attention control capability would retain less information. According to Fukuda et al., this was clear support for the attentional control account of individual differences in WMC estimates. This idea is consistent with the explanation from Unsworth and colleagues (e.g., Unsworth & Engle, 2007) that when an individual's memory capacity is exceeded, controlled processing must be initiated to manage the memoranda and retain as much information as possible. Individuals who engage this sort of processing will be able to perform close to their maximum capacity, whereas individuals who cannot engage in this sort of controlled processing will struggle to manage the additional information, become overloaded, and ultimately experience a cognitive breakdown of sorts. When this breakdown happens, an individual who normally has a capacity of, say, 3 items may only be able to maintain one of them. A similar idea was also expressed by Fukuda et al., who likened this phenomenon to an attentional overload.

However, it seems even clearer that selective visual arrays tasks place heavy demands on attention, to the extent that individual differences in selective visual arrays might be *primarily* due to attentional factors. For instance, Shipstead et al. (2014) found that two non-selective visual arrays tasks and two selective visual arrays tasks loaded well onto the same factor, but that the selective versions had a significant relationship to attention control above-and-beyond non-selective visual arrays and complex span tasks. Fukuda and Vogel (2009) showed that lower ability individuals orient their attention to distractors despite the cue, resulting in encoding and storing more information than necessary and thus lower *k* scores. Vogel et al. (2005) found that contralateral delay activity (which increases as more information is stored) for low and high WMC individuals was roughly the same for lower set size arrays when distractors were absent, but there were large differences in contralateral delay activity between high and low ability individuals for set sizes as low as 2 when two distractors were present. Finally, Robison et al. (2018) tested the contribution of WMC to filtering in visual arrays and found that WMC only explained 1-3% of the variance in selective visual arrays performance above and beyond non-selective visual arrays. I interpret this result to suggest that the filtering component of selective visual arrays places additional *attentional* demands on the respondent, as indicated by results of Shipstead, Vogel, and colleagues. To summarize, studies show that neurotypical individuals of differing ability can perform change detection tasks equally well so long as set sizes do not exceed capacity, but that either adding distractors or increasing set sizes to supra-capacity levels result in individual differences that are primarily attributable to attention control and not WMC.

Correlational results from our lab support the conclusions from Vogel and colleagues that distractors result in individual differences attributable to attention control. As shown by our reanalysis of four independent datasets from our lab described in Martin et al. (in press), non-selective and selective visual arrays cohere strongly but can generally be separated into two distinct, yet strongly correlated, latent variables. In exploratory factor analysis, non-selective visual arrays tasks load more strongly with other WMC tasks such as complex span and running span, whereas selective visual arrays loads more strongly with measures of attention control (namely antisaccade). In confirmatory factor analysis, when the selective visual arrays task was cross-loaded onto factors of both WMC and attention control, the general trend was for the loadings to the attention factor to be strong and loadings to the WMC to be weak, even non-significant. And model fit suffered immensely if selective visual arrays was forced to load solely with WMC as opposed to attention control. However, non-selective visual arrays showed more equal loadings to both WMC and attention control, with a tendency to load more with WMC. Given these results, it is not surprising that we found selective visual arrays (with set sizes of 5 and 7, thus supracapacity for most) to be such a strong measure of attention control in Draheim et al. (2020b).

In summary, while visual arrays clearly places demands on respondents to maintain information in primary memory, converging evidence suggests that *individual differences* in this paradigm are a strong reflection of differences in attention-related abilities. This is particularly true for the selective versions of the task in which distractors are present. The next section will be a discussion about how attention is allocated across the visual field, which is a diversion from the discussion regarding the measurement of attention control. However, subsequent sections will tie the two together.

1.3 Allocation of Attention Across the Visual Field

1.3.1 Types of Allocation and Individual Differences

Some proponents of early selection have likened visual allocation of attention to that of a movable spotlight in which attention is focused intensely on a small region of the visual field at the expense of peripheral information (LaBerge, 1983; Posner et al., 1980). Jonides (1983) argued for a two-process model in which visual attention resources can either be distributed over the entire visual field with equal attention and parallel processing of all items in the visual field, but with low resolution, *or* constricted to smaller portions of the visual field with high resolving power. Eriksen and Yeh (1985) argued instead that the ability to focus diffusely across the visual field or intensely on a particular location was on a continuum rather than discrete, more akin to the variable zoom of a lens. A low power setting of the zoom lens results in all areas of the visual field receiving an equivalent distribution of attentional resources, but this low density of resource allocation results in slow and limited processing for items within the field. As the zoom lens is powered up and attention is constrained to smaller regions of space, processing of information within this smaller space would become more rapid and the capacity to resolve finer detail or extract more information from the stimuli would improve. Importantly, the controlled act of focusing (or "zooming the lens") takes time. In a follow-up study, Eriksen and St. James (1986) found that the general tendency was for individuals to begin with attention in a diffuse state, attending to the entire visual field, but that participants used the focused spotlight approach in proportion to the validity of a pre-cue to a cued location. That is, the more likely a pre-cue was to validly indicate the position of target stimuli, the more constrained visual attention was for the cued region (i.e., smaller spotlight).

Of present interest is two sets of studies from our lab that explored individual differences in visual attention allocation: Heitz and Engle (2007) and Bleckley et al. (2003; 2014). Heitz and Engle administered Eriksen flanker trials in six blocks, with the first block having a 700 ms response deadline and each subsequent block having a deadline 100 ms quicker than the previous. They plotted performance using a conditional accuracy function (a plot of RT on the x-axis and accuracy on the y-axis) and found that although high and low WMC individuals reached the same level of asymptotic accuracy on mixed-block incongruent trials, high spans did so *quicker*. Heitz and Engle interpreted their results using the zoom-lens metaphor - high WMC individuals seemingly focused their lens quicker than low WMC individuals, but both groups had the same capacity to focus their lens if given enough time

In a series of experiments, Bleckley and colleagues examined the relationship between WMC and allocation of visual attention (Bleckley et al. 2003; 2014). They employed a selective attention task used by Egly and Homa (1984) in which participants attend to a central fixation surrounded by three concentric rings (octagons). The participants' goal was to identify a target letter in the center of the screen and also identify the location of a peripheral letter which appeared in one of eight locations on one of the three rings (24 total possible locations). Prior to stimulus display, participants either received (1) a valid cue, (2) an invalid cue, or (3) no cue as to the location of the peripheral target. If a cue was presented, 20% of the time it was invalid such that the peripheral target did not appear in the cued location (e.g., participants received the cue "MIDDLE" indicating they should expect the peripheral target to appear in the middle ring, but instead it appeared either in the outer or inner ring). The stimuli were displayed very briefly (around 40 ms) and then masked, with stimulation duration adaptive to each subject based on their performance in a practice session. Bleckley et al. (2003) found that high-WMC individuals were better able to identify the location of the peripheral letter than low-WMC individuals in the valid cue and no cue conditions when the peripheral letter was in either the middle or inner ring, but not the outer ring. Interestingly, performance of high-WMC participants also decreased substantially in invalidly cued trials, provided that the peripheral letter occurred closer to the center of the screen than the cued location (e.g., "DISTANT" was cued but the peripheral letter occurred in the inner ring), whereas low WMC participants' performance did not suffer on invalidly cued trials when the peripheral letter occurred closer to the center than expected. Furthermore, performance for low-WMC participants was better for invalidly cued trials when the peripheral letter appeared closer to the center than expected, as opposed to when the peripheral letter appeared farther from the center than expected (e.g., “CLOSE” was cued but the peripheral letter appeared in the outer ring). For high-WMC participants, performance on invalidly cued trials was not affected by whether the peripheral letter appeared closer or farther from the center than expected. From these results, Bleckley et al. inferred that low-WMC individuals were less flexible in their attention allocation. Specifically, low-WMC individuals tend to allocate attention in a spotlight manner, whereas high-WMC individuals allocate attention in a more object-based manner, allowing them to attend to discontiguous locations. This was interpreted as evidence that low-WMC individuals are also lower in attention control.

Bleckley et al. (2014) expanded on this finding by adding cognitive load (patterned finger-tapping) to the task in their first experiment and, in a second experiment, testing a different procedure designed to elaborate on the nature of differences in attentional allocation between high- and low-WMC individuals. A critical finding in their first experiment was that high-WMC participants appeared to utilize spotlight attention while under load, just as low-WMC participants use attention while not under load. While not under load, however, high-WMC participants showed more flexible attention allocation and were capable of discontiguous attention as in Bleckley et al. (2003). In their second experiment, Bleckley et al. (2014) employed a go/no-go procedure from Egly et al. (1994) which involved cueing participants to the probable (75%) target location at the end of one of two rectangles (bars) of equal size and orientation presented on either side of the screen. After a brief delay, one of the two rectangles had one of their ends highlighted, meaning there were a total of four possible locations for the highlighted (target) area. Respondents were asked to press a key as soon as one of the ends of the rectangles became highlighted, and do nothing if this did not occur (no-go trials). Three cue types were used: (1) valid cue in which the target area was contained by the cue, (2) invalid within-object in which the target area was within the same rectangle of the cue but on the opposite end, or (3) invalid between-objects, in which the target area was on the same side of the screen as the cued area but within the non-cued rectangle. RTs for validly cued trials were statistically quicker than RTs for trials with invalid between-objects cues, but high- and low-WMC individuals showed no overall differences in these RTs. Critically, however, low-WMC participants were equally slow to respond on trials with invalid cues, whereas high-WMC participants were quicker to respond to the trials involving invalid within-object cues. Several important findings came out of the results from Bleckley et al. (2003; 2014). First, individuals of higher cognitive ability are more able and/or likely to use object-based visual allocation than individuals of lower cognitive ability, who instead appear to rigidly use spotlight allocation. Second, individuals of higher cognitive ability are more flexible in their visual attention allocation and are capable of using spotlight-based allocation as well as object-based. Third, object-based allocation appears to be more demanding, as evidenced by high WMC individuals resorting to spotlight allocation while under cognitive load, whereas they use object-based allocation in the same task when not under load. This additional demand may be because object-based allocation requires more effortful and controlled processing, in other words greater attention control.

1.3.2 Limitations and Gaps in the Literature

The spotlight model of visual attention has been criticized for its simplicity and inability to account for (a) the dynamic nature of our environments, (b) discontiguous allocation of attention, and (c) object-based allocation (e.g., Bleckley et al., 2003; 2014; Cave & Bichot, 1999; Driver & Baylis, 1989; Müller et al., 2003; Valdes-Sosa et al., 1998). However, it is still a useful framework for behavioral phenomena in some situations and tasks (e.g., Eriksen & St. James, 1986), particularly with the flanker paradigm and/or conditions of high cognitive load (e.g., Bleckley et al., 2014; Eriksen & Eriksen, 1974; Heitz & Engle, 2007; Lavie, 1995). Given the findings discussed thus far, my view is that spotlight allocation is just one way in which individuals may allocate attentive resources across the visual field, and that there are a number of unexplored interesting questions concerning individual differences in spatial allocation of attention. First, several limitations of the Heitz and Engle (2007) and Bleckley et al. (2003; 2014) studies warrant mention. One limitation is that they were extreme group designs, which are resource efficient and useful for demonstrating the potential for an effect but suffer from a number of problems such as inflating statistical power, biasing effect sizes, lower reliability, misspecification, and regression to the mean (e.g., Preacher, 2015). Using quartile splits (which is quite common, and the approach taken by Heitz & Engle and Bleckley et al.), exacerbates the issues of inflated power and biased effect sizes (Feldt, 1961). Further issues are that extreme group designs are not optimal for precisely measuring the strength of a relationship, and that there are potential confounds. For instance, because of the strong relationship among WMC and Gf, individuals high in WMC are also highly likely to be high in Gf, and so it is not clear whether the results are due to differences in WMC or Gf.

As argued by Engle (2018), some previous studies by him and his colleagues which used high- and low-WMC groups and made attributions about WMC were likely flawed, and instead the results were due to high- and low-WMC individuals also being high and low in fluid intelligence or, as we now believe, *attention control*. Worse yet is that in all but one experiment of the seven total reported by Heitz and Engle and Bleckley et al., their quartile splits were not based on a robust assessment of WMC, but on performance from a single indicator of WMC – the operation span, which is not the psychometrically strongest measure of WMC (see Draheim et al., 2018). A more informative, albeit resource intensive, approach would therefore be to assess correlations across the entire ability range and to include multiple indicators of a variety of higher-order abilities to avoid some of these problems.

1.3.2.1 Donut-Shaped Allocation

More substantively, studies investigating the visual allocation of attention typically require participants to focally attend to the display, with distractors and/or additional targets occurring outside this focus. One question this raises is how attention is allocated when the *distractors* are in the center of the display, such that participants should allocate visual attention in a donut-shaped manner to ignore a focal distractor embedded within and surrounded by target stimuli. The only paper I could find on "donut" allocation was an electroencephalogram study by Müller and Hübner (2002) which provided some evidence that individuals are able to use this type of allocation. In another study, Beck and Lavie (2005) reported that when participants are asked to fixate on a location which will contain a distractor (as in donut-shaped displays), flanker interference effects are magnified. They concluded that distractors at fixation are therefore harder to ignore or filter out than peripheral distractors, perhaps because greater attentional weight is given to focal stimuli during response selection. Importantly, these two studies involved group-level analyses and were not designed to explore individual differences.

In summary, the literature supports that at least three different types of visual allocation of attention are utilized by individuals: object-based, spotlight, and donut-shaped.[3] Studies assessing individual differences in these sorts of allocation are sparse, limited in scope, and have notable methodological limitations. To that end, two primary questions of the present study are (a) to what extent are there individual differences in the ability to allocate visual attention in these various manners, and (b) do these individual differences have any predictive or theoretical value? More broadly, this gets at the largely unexplored area of to what degree are there important individual differences in strategic and flexible allocation of visual attention.

[3] Discontiguous allocation could be considered a fourth way in which individuals allocate attention across the visual field. But I consider discontiguous allocation of attention to be a result or product of donut or object-based allocation, rather than a separate type of allocation.

CHAPTER 2. THE PRESENT STUDY

The present student used the visual arrays paradigm to merge two areas of research: (1) the assessment of attention control and (2) the nature of allocation of attention in the visual field. Specifically, the goals are to explore individual differences in visual attention allocation more broadly and thoroughly than previous endeavors (e.g., Bleckley et al., 2003; 2014; Heitz & Engle, 2003) and to expand on the results of Draheim et al. (2021b) in continuance of the toolbox approach to assessing individual differences in attention control.

I argue that the visual arrays paradigm is well suited to accomplishing both these goals. The underlying assumption is that individual differences in change detection, specifically selective-based visual arrays, are primarily due to attention control, as argued by Draheim et al. (2021b) and Martin et al. (in press). Because the visual arrays paradigm is viewed as an estimator of memory capacity, only a handful of researchers have utilized it for the understanding of visual allocation of attention and instead such studies often use flanker, visual search, or simple target identification tasks. But because of the modular nature of visual arrays tasks and because it is clear that selective attention is a major source of individual differences in visual arrays performance, I argue that it is a viable paradigm of which to assess not only how visual attention is allocated, but how this allocation interacts with various factors such as cognitive ability, cognitive load, perceptual load, and preparation time. The analyses focus on several versions of visual arrays, however these tasks were embedded within an ongoing large-scale correlational study. Manipulations to visual arrays permitted testing some additional theoretical questions.

The following is a list of some of the questions of interest the present study was designed to address:

1. Can we replicate our previous findings that a) selective visual arrays is a strong measure of attention control, and b) attention control fully mediates the WMC-Gf relationship?

2. To what degree is attention control a separable and coherent construct? Is it multifaceted?

3. To what extent can individuals alter how they allocate attention in the visual field based on different task demands? Are higher ability individuals quicker to make the proper allocation of visual attention, better, or both? Does the type of allocation matter?

4. Do individual differences in above-capacity vs. near-capacity visual arrays reflect different processes? Are above-capacity demands required to maximize individual differences in these tasks?

2.1 Method

The study at large consisted of a battery of cognitive tasks administered over five two-hour long sessions. Included were measures of WMC, Gf, attention control (auditory and visual), processing speed, inspection time, multitasking, complex problem solving, and visual search. The emphasis for the present study is the attention control measures, particularly the visual arrays variants. The study at large is ongoing with a target sample size around N = 400, however for the purposes of fulfilling my dissertation requirement at the Georgia Institute of Technology, data analysis was cut off with N = 210 finishing the first three sessions. Note that tasks included in sessions 4 or 5 of the study at large will not be discussed in the method or results section, with the exception of the processing speed measures used for discriminant validity.

2.1.1 Participants and General Procedure

Data collection took place at the Georgia Institute of Technology. Participants were recruited from several colleges in the Atlanta, GA area as well as the general Atlanta community. To be eligible for the study, participants needed to report being a native (learned at age 5 or earlier) and fluent English speaker age 18 – 35 years with normal or corrected-to-normal vision, no history of seizure, and having never participated in a previous study in our lab.

Participants received \$200 compensation for their participation in the five-session study at large, distributed as \$30 for the first session with each subsequent session worth \$5 more than the previous. Due to COVID-19, participants were run in separate rooms, and our lab could accommodate up to three participants for sessions 1-4 at a time and an additional up to three for session 5. An experimenter was nearby, observing performance to ensure participants were following instructions. The experimenter also took notes regarding participant behavior, alertness, and apparent motivation, answered task-related questions, and started the run files for each task. Participants were not told directly that they would be observed but they were aware of the experimenter’s presence. Generally, undergraduate and post-baccalaureate students served as the experimenter, although graduate students and post-docs filled in as needed. At least one senior lab member (graduate student or post doc) was present to supervise data collection. Most tasks had built-in rest periods to appear after a set number of trials for each task, designed to occur approximately every ten minutes. Participants could advance the rest screen at their convenience to continue performing the task. Participants were asked to avoid getting out of their chair while in the middle of a task if possible and were encouraged to take short breaks between tasks when needed. The Georgia Institute of Technology Institutional Review Board approved the protocols (H20538 & H20532) for this study as well as several amendments throughout data collection. Finally, our lab followed Georgia Tech COVID19 guidelines and protocols, including COVID-19 screening for participants, temperature checks, mask-wearing, and physical distancing to the extent possible given lab layout. Lab personnel were also highly encouraged to get vaccinated and tested weekly. Participants were not allowed into the building if they answered "yes" to any question on the COVID19 checklist for possible symptoms and risk factors, and the lab was shut down on several occasions after a self-reported positive COVID-19 test from a participant or lab member.

2.1.2 Tasks of Interest

2.1.2.1 Task Order and Information

All WMC, Gf, processing speed, visual arrays, and other attention control tasks were programmed and run using E-Prime software. Multitasking measures were either in proprietary software or a standalone program. Tasks of the same construct were interspersed throughout the first four sessions of the study. That is, each session had a combination of attention, visual arrays, WMC, Gf, and multitasking tasks, along with various other tasks. One exception was the processing speed tasks, which all occurred during the 4th session. Several other tasks in the 4th session were omitted due to insufficient sample size ($n < 100$), which included running digit span (WMC), mental counters (WMC), paper folding (Gf), and control tower (multitasking). Novel or experimental attention control measures (e.g., auditory attention, double conflict versions of Stroop, flanker, and Simon) were also not analyzed in the present study. The full list and order of tasks in the first four sessions are presented in Table A1.

2.1.2.2 Working Memory Capacity

WMC was measured with two spatial complex span tasks (rotation and symmetry span), and a verbal running span task (running letter span).

The complex span tasks consist of alternating memory storage and processing subtasks (Unsworth et al., 2005; Figure 5). The advanced versions of the tasks included larger set sizes of memory items (Draheim, et al., 2018). For each task, to help ensure that participants were attending to the processing task and not rehearsing the to-be-remembered information during this time, participants were allotted up to their individual mean reaction time + 2.5 standard deviations from a block of practice trials on the processing task only. Additionally, participants were asked to maintain a minimum level of performance on the processing tasks (85% accuracy) and their cumulative percent correct was shown to them throughout the task in the upper-left corner of the screen.

Running span tasks involved administering a series of to-be-remembered material and then abruptly stopping the presentation and asking the participant to recall the last *n* stimuli in serial order.

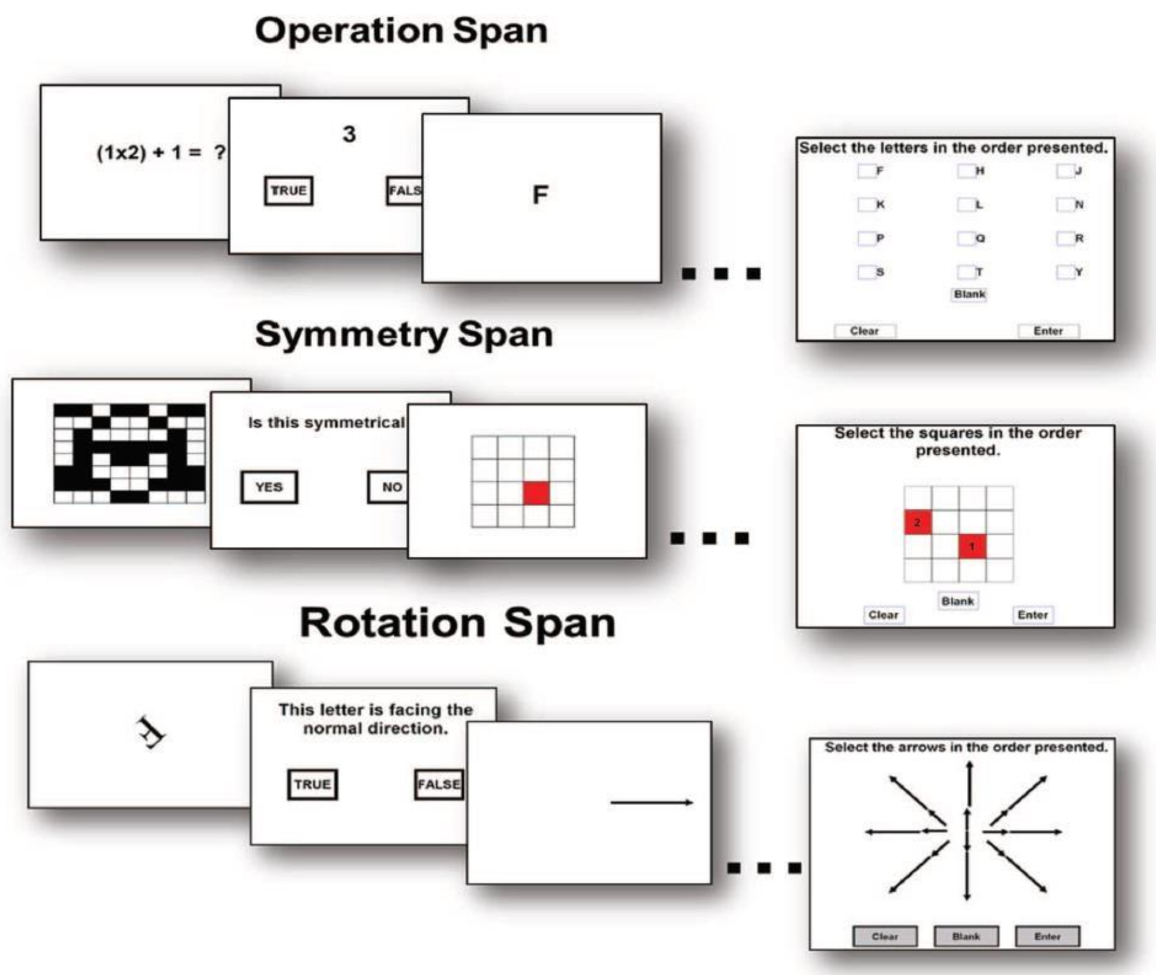

Figure 5 – The Complex Span Tasks. In each task, participants respond true/false or yes/no to a processing (distractor) task prior to the presentation of each to-be-remembered stimulus. After a variable number of presentations (depending on the set size for that trial), a recall screen appears asking the participant to recall the to-be-remember stimulus in order of presentation. The dependent variable is the partial span score, which is the total number of items recalled in the correct position. Note that the operation span is shown here but will not be included in the study.

2.1.2.3 Fluid Intelligence

Raven's advanced progressive matrices – Odd problems (Kane et al., 2004; Raven & Court, 1998). Participants were presented abstract shapes in a 3 x 3 matrix. The shape in the bottom-right was missing, and the participant was asked to select which item, from the eight answer options, best completed the overall pattern by clicking that option on the screen. Participants had 10 minutes to complete 18 problems. The number of correct responses was the dependent variable.

Number series (Unsworth et al., 2009; Thurstone, 1938). Participants were presented a sequence of numbers and needed to identify the response option that was the next logical number in the sequence by clicking the correct number from five total response options. Participants had 5 minutes to answer 15 problems. The number of correct responses was the dependent variable.

Letter sets (Ekstrom et al., 1976). On each problem, the participant was presented five sets of letters, each containing four letters that follow a particular rule. Instructions were to find the rule that applied to four of the five letter sets, and then indicate the set that violates the rule by clicking that set on the screen. Participants had 10 minutes to complete 30 problems. The number of correct responses was the dependent variable.

2.1.2.4 Attention Control (Non-Visual Arrays)

Antisaccade (Hutchison, 2007; Kane et al., 2001). Participants saw a central fixation cross lasting a random amount of time between 2000 – 3000 ms followed by an alerting tone for 300 ms. After the alerting tone, an asterisk appeared for 300 ms at 12.3° visual angle to the left or the right of the central fixation followed immediately by a target "Q" or an "O" for 100 ms on the *opposite* side of the screen from the asterisk. The location of the asterisk and target letter were then both masked for 500 ms by "##". The participant's goal was to ignore the asterisk and instead look away to the other side of the screen to catch the target "Q" or "O." Participants had as much time as needed to respond to which letter appeared by pressing the associated key on the keyboard. After responding, accuracy feedback was displayed for 500 ms, followed by a blank inter-trial interval of 1,000 ms. Participants completed 72 trials, and the dependent variable was the number of correctly identified target letters.

Flanker response deadline (flanker DL; Draheim et al., 2021b). This was a modified version of the arrow flanker involving an adaptive procedure to estimate each subject's response deadline threshold. In the basic arrow flanker paradigm, participants are presented with a target arrow in the center of the screen pointing left or right along with two flanking arrows on both sides. The four flanking arrows either point in the same direction as the central target (congruent trial; e.g., ← ← ← ← ←) or in the opposite direction (incongruent trial; e.g., ← ← ❼ ← ←). The participant is asked to indicate the direction the *central* arrow was pointing by pressing the "z" (left) or "/" (right) key. These keys have the words LEFT and RIGHT taped onto them to assist with response mapping.

In Draheim et al. (2021b), the adaptive flanker response deadline task was among the best performing attention measures according to our reliability and validity criteria, and was the top-performing adaptive measure. The version of this task used in Draheim et al. involved administering 18 blocks of 18 trials each (324 total) with a 2:1 congruent:incongruent ratio, and requiring the participant to be correct on at least 15 trials in each block in order for the response deadline to decrease on the subsequent block.

The version used in the present study was modified from that version to be adaptive on a trial-by-trial basis. Whenever an incongruent trial was presented, an incorrect response caused the response deadline to increase (more time to respond) on subsequent trials until the next incongruent trial. A correct response on an incongruent trial caused the response deadline to decrease (less time before a response was required). Piloting was conducted to help determine appropriate number of trials, starting point for the response deadline, and how much the deadline increases or decreases with each response. This task had a 2:1 congruent:incongruent ratio and a 3:1 up-down ratio so as to converge upon 75% accuracy (Kaernbach, 1991). The response deadline was implemented as a loud beep that indicated the participant has forfeited the opportunity to respond on that trial, and thus the trial was considered incorrect.

Stroop response deadline (Stroop DL; Draheim et al., 2021b). This was a modified version of the color Stroop involving an adaptive procedure to estimate each subject's response deadline threshold. In the basic Stroop paradigm, participants are presented with a color word (e.g., "RED") in colored font. The color of the font can match the word (congruent trial; e.g., **RED**) or can be a different color (incongruent trial; e.g., **RED**). The participant is asked to indicate the color of the font, therefore resisting the automaticity associated with reading the word, by pressing 1, 2, or 3 on the numpad. The response keys have green, blue, and red labels taped onto them to assist with response mapping.

In Draheim et al. (2021b), the adaptive Stroop deadline task was a clear improvement to the standard Stroop scored using RT and difference scores. The version of this task used in Draheim et al. involved administering 18 blocks of 18 trials each (324 total) with a 2:1 congruent:incongruent ratio, and requiring the participant to be correct on at least 15 trials in each block in order for the response deadline to decrease on the subsequent block.

The version used in the present study was modified from that version to be adaptive on a trial-by-trial basis. Whenever an incongruent trial was presented, an incorrect response caused the response deadline to increase (more time to respond) on subsequent trials until the next incongruent trial. A correct response on an incongruent trial caused the response deadline to decrease (less time before a response is required). Piloting was conducted to help determine appropriate number of trials, starting point for the response deadline, and how much the deadline increases or decreases with each response. This task had a 2:1 congruent to incongruent ratio and a 3:1 up-down ratio so as to converge upon 75% accuracy (Kaernbach, 1991). The response deadline was implemented as a loud beep that indicated the participant had forfeited the opportunity to respond on that trial, and thus the trial was considered incorrect.

Sustained attention-to-cue task (SACT; Draheim et al., 2021b; Figure 6). This task was designed as an accuracy-based analog of the psychomotor vigilance task. In Draheim et al. (2021b), it was the third-best performing measure of attention control behind visual arrays and antisaccade according to our reliability and validity criteria.

In this task, participants need to sustain their attention on a visual circle cue presented at random locations on the screen and ultimately identify a target letter presented briefly at the center of the cue. The stimuli for the task were presented against a grey background. Each trial started with a central black fixation. On half of the trials, the fixation was presented for 2 seconds and for the other half the fixation was presented for 3 seconds. After the fixation, following a 300 ms tone, a large white circle cue was presented in a randomly determined location on either the left or right side of the screen. To orient the participant on the circle cue, the large circle began to immediately shrink in size until it reaches a fixed size. Then, a 3 x 3 array of letters was displayed at the center of the cue. The letters in the array consisted of B, D, P, and R. The central letter was the target letter and was presented in a dark grey font. The non-target letters were presented in black font with each letter occurring twice in the array and the target letter occurring three times. After 125 ms the central letter was masked with a "#" for 1000 ms. After the mask, the response options were displayed in boxes horizontally across the upper half of the screen. The participant used the mouse to select whether the target was a B, D, P, or R. Feedback was given during the practice trials but not the experimental trials. Sixty-four trials were administered. Accuracy rate was the dependent variable.

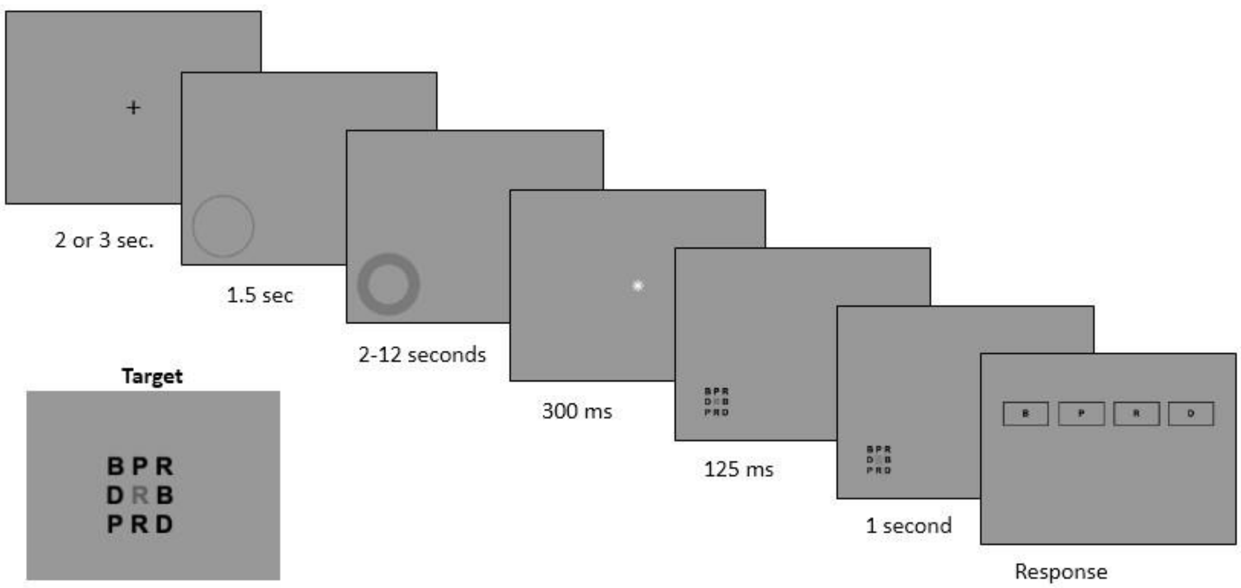

Figure 6 – Trial sequence for the sustained attention-to-cue task. Participants saw a fixation for 2 or 3 seconds followed by a circle cue indicating the future location of the target letter array. This circle shrank for 1.5 seconds and then remained for either 2, 4, 8, or 12 seconds. Then, the target 3 x 3 letter array (see enlarged view at bottom left corner) appeared for 125 ms and was masked for 1000 ms. Participants indicated which of four possible letters (B, D, P, or R) appeared in the center.

2.1.2.5 Attention Control – Visual Arrays (Luck & Vogel, 1997; Shipstead et al., 2014)

Each visual arrays task involved set sizes 3 (near-capacity for the typical participant) and set sizes 5 (above capacity for most participants). The rationale for including a near-capacity and above-capacity set size was based on research showing that individual differences in visual arrays are due, in part, to whether set sizes exceed capacity limits of the individual (e.g., Fukuda et al., 2015; Vogel et al., 2005). The goal was therefore to determine whether supracapacity set sizes are necessary to find strong individual differences in these measures. Further, it is possible that cognitive and perceptual load (indexed by set size) could interact with how attention is allocated (e.g., Lavie, 1995).

The visual arrays tasks had 60 trials of each set size (3 and 5) except the enhanced concentric visual arrays (described below) which had 54 in order to equally balance the number of trials for each trial type.

Each visual arrays task had a 500 ms intertrial interval, which is not depicted in the task figures below. **Participants received no feedback regarding their performance in the experimental trials.**

Performing the visual arrays tasks required a binary yes or no response from the participant as to whether the response display is the same as the initial display. Participants responded by pressing to-be-determined keys labeled “Yes” and “No.” To assist with response mapping, the “Yes” response corresponded to the S key (for same) and the “No” response was the D key (for different) on the keyboard. For consistency, in each visual arrays task all potential targets from the initial array were presented in the response (probe) array, but one target was probed with a white dot such that the respondent must judge whether that particular rectangle changed in orientation from the initial array.

The dependent variable for visual arrays performance is a capacity score (*k*) calculated using the single probe correction (see Cowan et al., 2005; Shipstead et al., 2014). This calculation is N * (hits + correction rejections – 1), where N is the set size for that array. This calculation is done separately for each set size. *K* scores can be converted into accuracy scores by first adding the set size from which the *k* score was calculated and then dividing by 2x the set size.

Non-selective visual arrays (NSVA; Figure 7). Non-selective visual arrays was included as a baseline to allow comparison to the selective versions. Participants saw an array of either 3 or 5 blue rectangles, each of which was in a random orientation of four possible orientations. The array was presented for 250 ms and, after a delay of 900 ms, the array was presented again and one of the target rectangles had changed orientation from the original array on 50% of the trials and was probed with a white dot. The participant was asked whether the probed rectangle was in the same orientation as it was in the initial array.

Participants performed 12 trials of practice on set sizes 1-3 in which the initial array had a presentation rate 4x slower than the experimental trials and with trial-by-trial feedback. Each participant had to achieve at least 75% accuracy to advance to the 2nd stage of practice. If they did not, they received a message indicating that they got too many practice trials wrong which explained the instructions again before allowing them to retry the slow practice trials. In the 2nd stage of practice, participants performed 12 trials that were identical to the experimental ones in timings and set sizes but with trial-by-trial feedback. Regardless of performance, after the 2nd stage of practice the 120 experimental trials were presented.

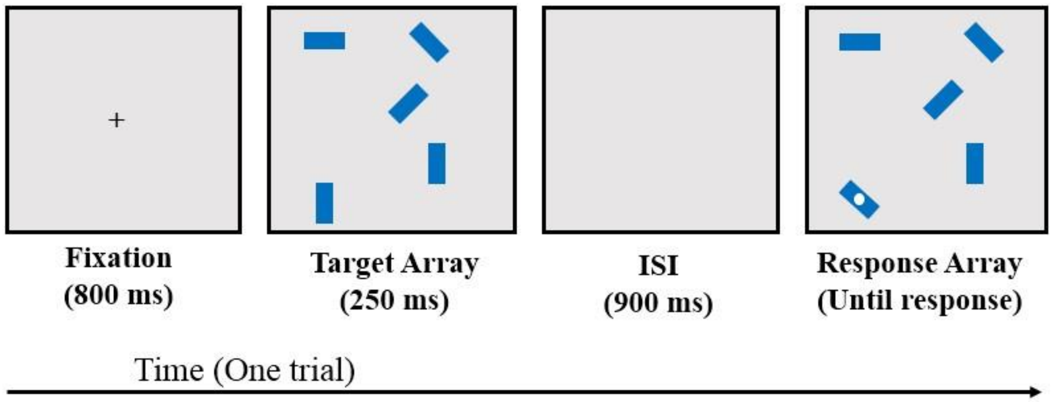

Figure 7 – Non-Selective Visual Arrays (NSVA). *Note*. Participants saw three or five blue rectangles. After the target array and a 900 ms-ISI, the rectangles were re-presented, one of which had a white dot on it. The participant's goal was to indicate whether this probed rectangle was the same orientation as in the initial array, and the participant was asked to respond "Yes" if the rectangle was in the same orientation and "No" if the rectangle had changed. In the trial shown, the participant should indicate "No". Example shown is set size 3. Not to scale.

Color selective visual arrays (VA4; Figure 8). Participants saw an array of an equal number of blue and red rectangles differing in orientation. Prior to each trial, the participant was cued to attend to either the red or blue rectangles by a 300 ms flash of *RED* or *BLUE.* Immediately after cue, the array was presented for 250 ms and, after a delay of 900 ms, the array was presented again with only the target rectangles. One of these rectangles changed orientation from the original array on 50% of the trials and was probed with a white dot. The participant was asked whether the probed rectangle was in the same orientation as it was in the initial array.

Participants performed 12 trials of practice on set sizes 1-3 in which the initial array had a presentation rate 4x slower than the experimental trials and with trial-by-trial feedback. Each participant had to achieve at least 75% accuracy to advance to the 2nd stage of practice. If they did not, they received a message indicating that they got too many practice trials wrong which explained the instructions again before allowing them to retry the slow practice trials. In the 2nd stage of practice, participants performed 12 trials that are identical to the experimental ones in timings and set sizes but with trial-by-trial feedback. Regardless of performance, after the 2nd stage of practice the 120 experimental trials were presented.

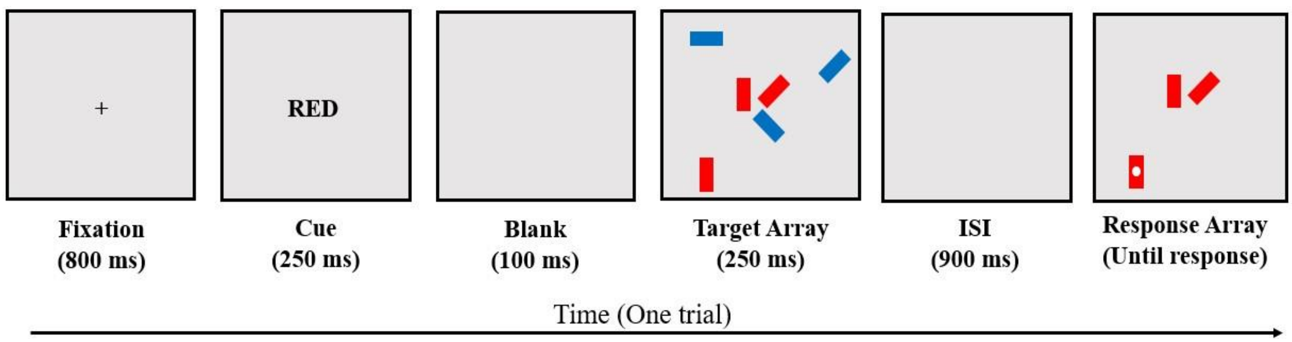

Figure 8 – Color Selective Visual Arrays (VA4). *Note.* Participants saw six or ten total rectangles, half blue and half red. The participant was cued to only attend to rectangles of a particular color. After the target array and a 900 ms-ISI, the rectangles of the cued color were re-presented, one of which had a white dot on it. The participant's goal was to indicate whether this probed rectangle is the same orientation as in the initial array, and the participant was asked to respond "Yes" if the rectangle was the same orientation and "No" if the rectangle had changed. In the trial shown, the participant should indicate "Yes". Example shown is set size 3. Not to scale.

Enhanced color selective visual arrays (EVA4; Figure 9). Prior research has found that individual differences in visual arrays are more pronounced when distractors are added, and that lower ability individuals cannot effectively suppress or otherwise filter out these distractors (Cowan & Morey, 2006; Fukuda & Vogel, 2009; Vogel et al., 2005). Therefore, increasing the number of distractors should magnify attentional-related individual differences.

Participants saw an array with an equal number of blue, red, and purple rectangles differing in orientation. Prior to each trial, the participant was cued to attend to which color of rectangles to attend to by a 300 ms flash of *RED*, *BLUE*, or *PURPLE.* Next, the array was presented for 250 ms after a delay of 900 ms, and then the array was presented again with only the target rectangles (red, blue, or purple). One of these rectangles had changed orientation from the original array on 50% of the trials and was probed with a white dot. The participant was asked whether the probed rectangle was in the same orientation as it was in the initial array.

Participants performed 12 trials of practice on set sizes 1-3 in which the initial array had a presentation rate 4x slower than the experimental trials and with trial-by-trial feedback. Each participant had to achieve at least 75% accuracy to advance to the 2^{nd} stage of practice. If they did not, they received a message indicating that they got too many practice trials wrong which explained the instructions again before allowing them to retry the slow practice trials. In the 2^{nd} stage of practice, participants performed 12 trials that were identical to the experimental ones in timings and set sizes but with trial-by-trial feedback. Regardless of performance, after the 2^{nd} stage of practice the 120 experimental trials were presented.

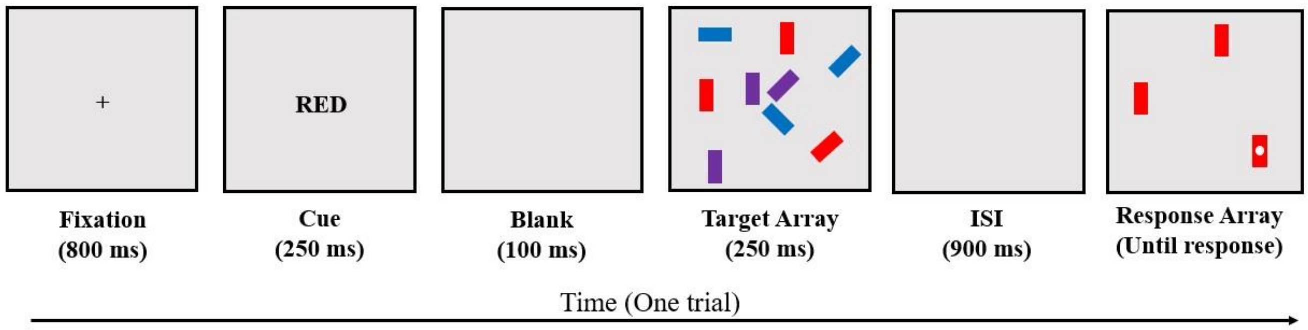

Figure 9 – Enhanced Color Selective Visual Arrays (EVA4). *Note.* Participants saw nine or fifteen total rectangles, 1/3 red, 1/3 blue, and 1/3 purple. The participant was cued to only attend to rectangles of a particular color. After the target array and a 900 ms-ISI, the rectangles of the cued color were re-presented, one of which had a white dot on it. The participant's goal was to indicate whether this probed rectangle was the same orientation as in the initial array, and the participant was asked to respond "Yes" if the rectangle was the same orientation and "No" if the rectangle had changed. In the trial shown, the participant should indicate "No". Example shown is set size 3. Not to scale.

Concentric selective visual arrays (CVA; Figure 10). It has been reported that selection demands in visual arrays based on location is considerably easier than selection by color (e.g., Anllo-Vento & Hillyard, 1996; Vogel et al., 2005). However, these designs generally involved dividing the display into hemifields or quadrants and cueing the participant to attend to the targets in one these areas. Even with this paradigm, individual differences still emerge with as few as two stimuli per quadrant (two target stimuli, six distractors). I proposed that perhaps an improvement to the standard location-based visual arrays would be to have target stimuli either inside or outside of a central focal location instead of using a quadrant or hemifield system. Additionally, such a task permits testing differences in two types of visual allocation – spotlight (focal target with peripheral distractors) and donut (focal distractors with peripheral targets).

Unlike the previously described visual arrays tasks, in the concentric version the delay between cue and target array had one of three possible durations (100 ms, 400 ms, or

700 ms). The reason for this variable cue-to-target duration was to assess potential individual differences in preparatory time to constrain attention, similar to the Heitz and Engle (2007) rationale for including increasingly shorter response deadlines in their flanker task. Heitz and Engle found that high and low-WMC individuals could constrain attention to a similar degree if given enough time, but that high spans were quicker to do so. Because of the nature of visual arrays, instead of response deadlines, preparatory cues would conceivably serve the same purpose – with longer cue-to-target times affording subjects more time to endogenously constrain their attention to the appropriately sized spotlight or donut, depending on which type of allocation was cued. The differences in high and low spans manifested with response deadlines of around 400 – 600 ms, however that was with a different paradigm. Because this a novel manipulation and variant of visual arrays, the cue-to-target durations covered a wider breadth.

Participants saw an array of 6 or 10 blue rectangles in various orientations, with half inside a small (3° visual angle) circle at the center of the screen and half along a ring with 9.3° visual angle (rectangles could appear in one of 12 equally-spaced locations on this ring, corresponding to hours on a clock). Prior to each trial, the subject was cued to attend to rectangles that were inside the focal circle or on the outside ring, by a 250 ms flash of either "INSIDE" or "OUTSIDE." Immediately after cue, the array was presented for 300 ms and, after an ISI of 900 ms, the array was presented again with only the rectangles in the cued region. One of these rectangles had changed orientation from the original array on 50% of the trials and was probed with a white dot. The participant was asked whether the probed rectangle was in the same orientation as it was in the initial array.

Due to the increased complexity of this task, practice was more involved and took place over four stages. In the first stage, participants performed 10 trials with a pictorial and spatial cue that showed the region they should attend to in light orange with the word "INSIDE or OUTSIDE" on it (see Figure 11) followed by 1-3 stimuli in that shaded area and 1-3 stimuli outside of it, and with a slow cue duration (1000 ms) and target duration (2500 ms). Subjects were asked to simply <u>COUNT</u> the number of stimuli in the area of interest to ensure they understand the purpose of the cue and could distinguish stimuli that are inside or outside the center. Participants received trial-by-trial accuracy feedback and had to respond correctly on 90% of these counting trials to advance to stage 2 of practice. If they responded incorrectly on more than one of these ten trials, they were shown a screen informing them that they were incorrect on too many trials which also reminded them of the instructions and instructed them to alert the experimenter if they did not understand these instructions. Then, they received the full instructions again and re-performed the ten counting trials. If a participant failed this stage of practice three times, then still advanced to the 2nd stage of practice.

In the 2nd stage of practice, cue duration was set to 500 ms, cue-to-target interval was 500 ms, and target duration was set to 1000 ms. Participants performed 12 trials with the spatial and verbal cue as with the 1st stage of practice and of set size 1-3, but in this stage their job was to perform the task as normal – indicating whether or not a target rectangle had changed orientation. Trial-by-trial accuracy feedback was provided, and participants had to respond correctly on at least 75% of these trials to advance to the 3rd stage of practice. If they did not, they again saw a screen informing them that they were incorrect on too many trials, reiterating the instructions, and asking them to alert the experimenter if they did not understand the instructions. Then they receive instructions again and performed the 12 practice trials. This process repeated until they reached 75% or above accuracy.

The 3rd stage of practice was identical to the 2nd, except the spatial cue was removed. Participants were only cued with the words "INSIDE" or "OUTSIDE" as to which area of the screen the target rectangles would appear.

The 4th stage of practice involved 12 practice trials that were identical to the experimental trials in set size and target duration, but with a static 400 ms cue-to-target delay instead of the variable delay of 100, 400, or 700 ms. Further, subjects received trial-by-trial accuracy feedback unlike the experimental trials. There was no practice-to-criterion for this stage of practice trials, all participants advanced to the experimental trials after performing this stage once.

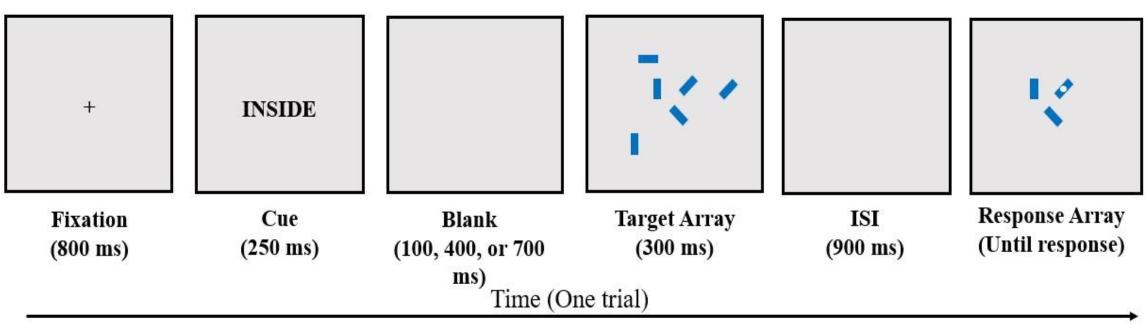

Figure 10 – Concentric Selective Visual Arrays (CVA). *Note.* Participants saw six or ten blue rectangles, half of which were presented focally and half on an imaginary ring 9.3° from the center. The participant was cued to only attend to rectangles either at the center or on the periphery by cues of "INSIDE" or "OUTSIDE". After the target array and a 900 ms-ISI, the rectangles in the cued location were re-presented, one of which had a white dot on it. The participant's goal is to indicate whether this probed rectangle was the same orientation as in the initial array, and the participant was asked to respond "Yes" if the rectangle was the same orientation and "No" if the rectangle had changed. In the trial shown, the participant should indicate "Yes". Example shown is set size 3. Not to scale.

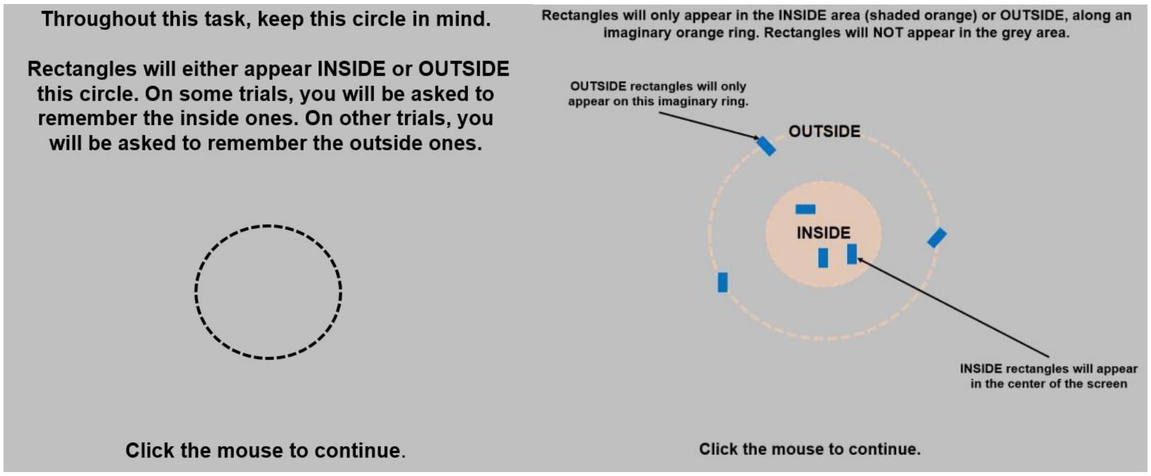

Figure 11 – Practice Figures for the Concentric Selective Visual Arrays Task. *Note*. Participants were given additional practice in the concentric tasks. The above images show two of the slides that they saw during instructions. Further, during practice the target location was cued both with the word "INSIDE" or "OUTSIDE" and with an orange indicator over the exact target area. The orange indicator in the cue was removed in the last block of practice.

Enhanced concentric selective visual arrays (ECVA; Figure 12). This task was a combination of the enhanced color selective visual arrays (2:1 distractor to target ratio) and the concentric visual arrays. Stimuli appeared in one of three regions – an inner circle (within 3° visual angle), a ring of 9.3° visual angle (rectangles could appear in one of 12 equally-spaced locations on this ring, corresponding to hours on a clock), and a ring of 18.5° visual angle (rectangles could appear in one of 24 locations on this ring with possible locations equally spaced by 15°). Just as with the concentric version with two regions of interest, this task involved a variable cue-to-target interval lasting 100, 400, or 700 ms.

Participants saw an array of 9 or 15 blue rectangles in various orientations, with 1/3 inside a small (3° visual angle) circle at the center of the screen, 1/3 along a ring with 9.3° visual angle, and 1/3 along a ring with 18.5° visual angle. Prior to each trial, the participant was cued to attend to rectangles that were inside the focal area, on the middle ring, or on the outer ring by a 250 ms flash of "INSIDE", "MIDDLE", or "OUTSIDE." Immediately after cue, the array was presented for 300 ms and, after a delay of 900 ms, the array was presented again with only the rectangles in the cued region. One of these rectangles had changed orientation from the original array on 50% of the trials and was probed with a white dot. The participant was asked whether the probed rectangle was in the same orientation as it was in the initial array.

Due to the increased complexity of this task, practice was more involved. See the above description of the practice procedure for the other concentric visual arrays task and also see Figure 13.

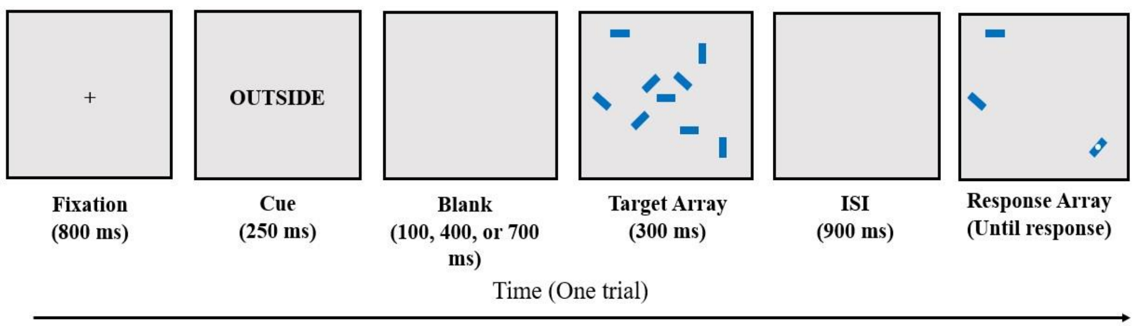

Figure 12 – Enhanced Concentric Visual Arrays Task (ECVA). *Note.* Participants saw ten or fifteen blue rectangles, 1/3 of which presented at the center of the screen, 1/3 on an imaginary ring 9.3° from the center, and 1/3 on an imaginary ring 18.5° from the center. The participant was cued to only attend to rectangles in the target area via a cue of "INSIDE", "MIDDLE", or "OUTSIDE". After the target array and a 900 ms-ISI, the rectangles in the cued location were re-presented, one of which had a white dot on it. The participant's goal is to indicate whether this probed rectangle was the same orientation as in the initial array, and the participant was asked to respond "Yes" if the rectangle was the same orientation and "No" if the rectangle had changed. In the trial shown, the participant should indicate "No". Example shown is set size 3. Not to scale.

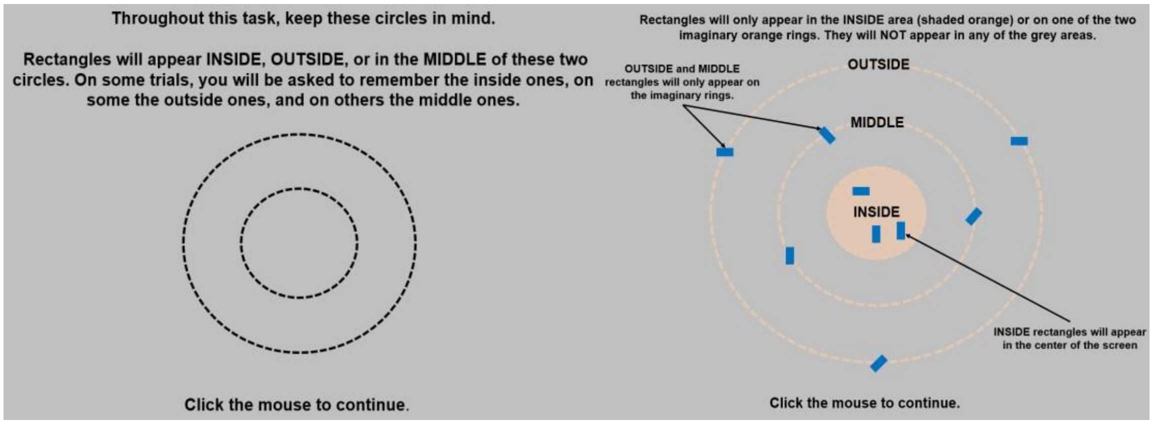

Figure 13 – Practice Figures for the Enhanced Concentric Selective Visual Arrays Task. *Note*. Participants were given additional practice in the concentric tasks. The above images show two of the slides that they saw during instructions. Further, during practice the target location is cued both with the word "INSIDE", "MIDDLE", or "OUTSIDE" and with an orange indicator over the exact target area. The orange indicator was removed in the last block of practice.

2.1.2.6 Multitasking

Multitasking tasks are often used as a proxy of real-world performance, and so two multitasking measures were included here as a criterion measure to assess predictive validity.

Synthetic work (SynWin; Elsmore, 1994). Synthetic work is a proprietary multitasking measure that requires concurrent processing of four independent tasks, both auditory and visual in nature. Participants performed three 5-minute blocks of trials, and scores were calculated using a formula that combined points earned across each subtest. Subtasks were basic arithmetic, probed memory recognition, visual monitoring of a fuel gauge, and auditory monitoring for an infrequent tone.

Foster task (Martin et al., 2020). Foster task was modeled after the synthetic work task and also involves participants performing four concurrent visual subtasks. Participants also performed three 5-minute blocks. Performance was scored as the total score across the four subtasks, averaged across the three blocks. Participants received 100 points for each correct response within a brief time window, -100 points for an incorrect response within that window, and also when participants failed to respond when required, their total score would begin to rapidly fall until either a response was made or their total reached 0. Subtasks were simple arithmetic, monitoring of a visual disc, word recall, and telling time on an analog clock.

2.1.2.7 Processing Speed

Processing speed measures were computerized versions of paper and pencil tests. In each case, subjects were instructed to respond as quickly and accurately as possible, but consistent with standard administration procedures, were not alerted of the time limits of each task in the instruction phase. Note that processing speed appeared in session 4 of the study at large, meaning relatively few participants completed this task compared to the other tasks included in the present study (listwise $N = 97$). As such, the processing speed measures were used sparingly in analyses.

Letter string comparison (Conway et al., 2002). Participants viewed strings of three, six, or nine consonants appearing to the left and right of a central line. The letter strings could either be the same or differ by a single letter. If different, the mismatching letter can appear in any location in the string. Responses were made by clicking on a button on the screen labeled SAME for identical strings or DIFF for mismatching strings. Letters were printed in white size 18 Courier New font on a black background. After completing six practice trials, two 30-second blocks of the task were administered. The dependent variable is the number of accurate responses across both blocks.

Digit string comparison. This task is identical to letter string comparison, except the participant was shown strings containing digits.

Pattern string comparison. This task is identical to letter string comparison, except that patterns were shown instead of letters.

2.1.3 Data Preparation

Data processing was kept to a minimum to maintain data purity. Missing data can also occur for a variety of other reasons including lost data file, experimenter error, or software/hardware error. In total, 11 of 210 participants had a single task score missing for any of the visual arrays measures, meaning 1.04% of scores for the visual arrays tasks were missing and 199 participants had scores for all visual arrays tasks. A t-test showed that these 11 participants had slightly larger scores on the color visual arrays task (which had scores for all 210 participants), but this was not statistically significant (two-tailed p = .355).

Outliers were checked by manual inspection and with some assistance from graphs. I determined that there were relatively few outliers in the dataset, and that the data were remarkably clean in this regard, particularly with the visual arrays tasks (see Figure 13 for distribution and Table 1 for descriptive statistics). For example, negative k scores (below chance performance) are common in visual arrays tasks, but only seven of the total scores across all the visual arrays tasks (0.67%) were negative, and with one exception noted below the average magnitude of the negative *k* scores was just -.08. Further, only five individual scores for any of the visual arrays tasks were below a z-score of -3.0 (none were above 3.0), two for VA4, two for EVA4, and one for CVA. Removing these scores resulted in no-to-minimal changes with correlations to other attention measures. As such, negative *k* scores and low z-scores were left as is, with the exception noted below.

One participant had a highly negative *k* score on the non-selective visual arrays task (-2.08) but highly positive Z scores on most other cognitive tasks, including the other visual arrays tasks. Given that the non-selective visual arrays is the first task of the study, it is almost certain that this participant mixed up the response mapping throughout the task. As such, their response for the non-selective visual arrays task were reversed, which produced scores in line with their scores on all other tasks. Additionally, one participant had a .39 score on antisaccade (well below chance) despite positive Z scores on the other cognitive tasks. It is not clear how one could confuse the response mapping or instructions in this task, but it was the only other clear outlier and so that data point was removed.

Missing scores were more common in some other tasks, with an average of 2.44% of data points missing in the other attention control, WMC, and Gf tasks: Antisaccade = 5.7%, SACT = 2.9%, Stroop DL = 0.5%, flanker DL = 0%, Raven = 1.9%, letter sets = 0.5%, number series = 0%, symmetry span = 4.3%, rotation span = 2.9%, and running letter span = 5.7%. Finally, the multitasking measures had 184 valid listwise datapoints (synthetic work = 6.2% missing, Foster task = 6.7% missing), and processing speed measures from session 4 had a listwise *N* of just 97. I decided to not remove any participants from analyses because the majority of participants with a data point missing on these tasks had no missing data on the visual arrays tasks, which were the most important of the present study. Further, many analyses involve using factor scores, and so a participant with any missing data will automatically not be included in those analyses as complete data for each construct in order to estimate factor scores. Data imputation was considered, but I decided against that in service of the larger goal to keep the dataset as veridical as possible.

Data were aggregated using E-Merge (a function of E-Prime), R scripts, and SPSS. Analyses were primarily conducted using SPSS; confirmatory factor analysis and structural equation modeling were performed in R.

2.2 Results

2.2.1 Task Abbreviations

Note that the following abbreviations were used in reporting results. NSVA = nonselective visual arrays, VA4 or VA = color selective visual arrays, EVA4 or EVA = enhanced color selective visual arrays, CVA = concentric visual arrays, ECVA = enhanced concentric visual arrays, VA-S = selective visual arrays (to distinguish the selective tasks from the non-selective task), SACT = sustained attention-to-cue, RAPM = Raven's advanced progressive matrices, RotSpan = rotation span, SymSpan = symmetry span.

2.2.2 Piloting

Extensive piloting work was done on a number of the tasks to ensure that participants understood the task and I made adjustments to task parameters and characteristics as needed. Of present concern is the extent to which participants could distinguish which spatial location the stimuli were in in the concentric tasks. In other words, did participants know which stimuli they were supposed to attend to? Initial piloting work suggested that participants were relatively poor at this, at which point I reprogrammed the concentric tasks such that the rectangles that were not in the center could only appear in certain locations along a ring equidistant from the center of the screen (i.e., in a ring around the center), instead of the initial versions in which targets would appear randomly so long as they were far enough from the center. I also made the rings larger than in initial versions, to add more distance between stimuli from the different regions.

The concentric tasks required participants to perform a “counting” task in which they were asked to count the number of stimuli only in a target region. For example, in the ECVA task they saw the cue “OUTSIDE” which informed them to only count the rectangles on the periphery of the screen (18.5° from the center). After ten trials of this, participants were only permitted to advance to the next stage of practice if they had responded accurately to at least 9 of those 10 counting trials. If they did not, they were sent back to the beginning of the task to redo the practice up to that point. But if they failed three total times, they were permitted to advance to the next stage of practice.

The tasks tracked the number of times participants failed practice. On the CVA task, 94.7% of participants did not fail the first counting practice, 3.9% failed exactly once, 1.4% failed exactly twice, and no participants failed three times. On the ECVA task, 93.8% did not fail the counting practice, 3.8% failed exactly once, 1.9% failed exactly twice, and .5% (one person) failed three times before being advanced to the next stage of practice. The number of people who failed a counting practice is too low to conduct significance testing as to whether they performed worse on the tasks as well. However, looking at the mean differences in performance on the tasks, participants who failed the CVA counting practice once performed very similarly to those who did not, and participants who failed the twice actually had a higher *k* score on the task than people who did not fail the counting practice. In the ECVA, people who failed the counting practice once had a .29 lower *k* score than those who did not fail, and people who failed two or three times had a much lower *k* score. Still, the vast majority of participants did not fail the counting practice, and only one participant on one of the tasks failed three times (that is, they never successfully completed it). I take this as evidence that participants generally were aware which rectangles belonged to which regions of interest in the tasks, and did not remove any scores from the concentric tasks on the basis of failing the counting practice.

2.2.3 Reliability

I calculated a lower-bound estimate of split-half reliability for each visual arrays task by correlating overall performance on set size 3 with set size 5, then applying the standard Spearman-Brown prophecy formula as a correction. Despite this method being a lower-bound, the estimates were fairly high and indicated adequate reliability: nonselective visual arrays had a split-half reliability of .78 and the other four visual arrays tasks ranked from .83 to .86. We can also see from subsequence analyses that, given the magnitude of correlations involving these tasks, reliability was not a concern for the visual arrays tasks.

2.2.4 Descriptive Statistics

Descriptives are listed in Table 1. Unexpectedly, the non-selective visual arrays was the easiest task, with an average *k* score of 2.74 (84.3% accuracy) and the enhanced concentric visual arrays was the most difficult, with an average *k* of 1.91 (73.9% accuracy). The other three tasks fell in-between, with the enhanced color visual arrays the most difficult among them (*k* = 2.3, 78.8% accuracy). Concentric visual arrays had the largest standard deviation and negative skew, whereas non-selective visual arrays had the most kurtosis. Overall, though, performance was quite normally distributed for behavioral data (Figure 14).

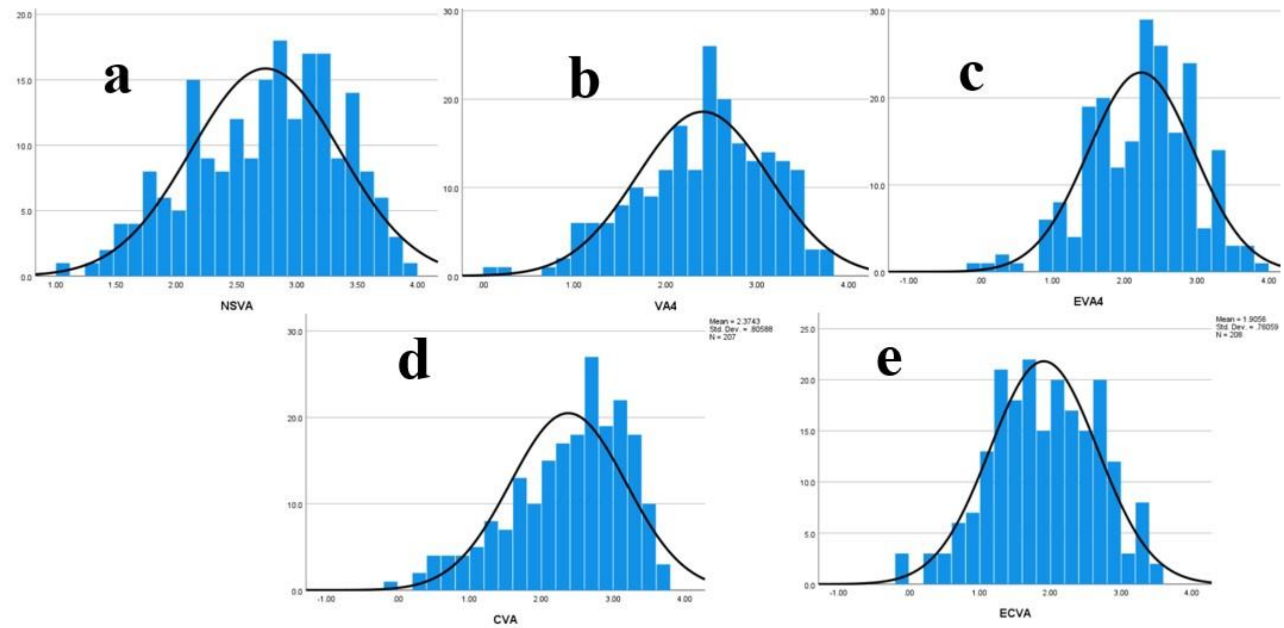

Figure 14 – Histograms of Visual Arrays Performance. Note. (a) Non-selective visual arrays; (b) Color selective visual arrays; (c) Enhanced color selection visual arrays; (d) Concentric visual arrays; (e) Enhanced concentric visual arrays.

Table 1 – Descriptive Statistics for the Visual Arrays Tasks.

Descriptive Statistics

	N	Range	Minimum	Maximum	Mean	Std. Deviation	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
NSVAScore	204	2.90	1.05	3.95	2.7382	.61575	-.344	.170	-.623	.339
VA4Score	210	3.70	.08	3.78	2.4053	.72156	-.487	.168	-.058	.334
EVA4Score	210	3.95	-.12	3.83	2.2250	.73104	-.437	.168	.186	.334
CVAScore	207	3.68	-.07	3.62	2.3743	.80588	-.744	.169	-.057	.337
ECVAScore	208	3.59	-.13	3.46	1.9056	.76059	-.175	.169	-.364	.336
Valid N (listwise)	199									

Note. The dependent variable for the visual arrays tasks is a capacity score (*k*) calculated using the single probe correction (see Cowan et al., 2005; Shipstead et al., 2014). This calculation is N * (hits + correction rejections – 1), where N is the set size for that array. This calculation is done separately for each set size.

2.2.5 Intercorrelation and Factor Structure

Intercorrelations among the visual arrays measures was quite strong, ranging from r = .55 - .80 with an average of r = .69 or 48% variance at the task level. Interestingly, the enhanced concentric version had the lowest average correlation with all other visual arrays tasks (r = .64), just behind non-selective visual arrays (r = .65). The average correlation for the other three tasks was r = .72 - .73. Correlations between VA4, EVA4, and CVA were quite strong.

Table 2 – Descriptive Statistics for the Visual Arrays Tasks.

Task	NSVA	VA4	EVA4	CVA	ECVA
Non-Selective	*.78*				
Color Selective	.72	*.84*			
Enhanced Color Selective	.66	.74	*.83*		
Concentric	.66	.80	.78	*.85*	
Enhanced Concentric	.55	.65	.68	.68	*.86*

Note. All correlations are statistically significant. Diagonals are estimated split-half reliability.

I performed a series of exploratory factor analyses using the attention control, WMC, and Gf tasks to determine the underlying factor structure of the data (Figure 15; Figure 16). This was done as a first pass to ensure tasks were behaving as expected, but also to see whether the visual arrays tasks would load with the other attention measures, with WMC, or on their own. All factor analyses were done using a direct oblimin (oblique) rotation with delta set at the default 0, and factors with eigenvalues greater than 1 were retained. Direct oblimin rotation results in two outputs that could be considered factor loadings: the rotated pattern matrix contains the standardized regression coefficients of the indicator onto the factors and the rotated structure matrix contains the zero-order correlations between indicators and factors. The pattern matrix is usually reported because it is easier to interpret in that it maximizes the differences between the factors. As such, pattern matrices are reported below, though I note that interpretation of the factors were the same whether pattern matrices or structure matrices were assessed.

For analyses in which three factors were extracted, these three factors generally accounted for around 60-65% of the variance, with factor 1 ~40%, factor 2 ~12%, and factor 3 ~10%. The order of the factors depended on which tasks were included. Factors were correlated around the r = .40 - .45 range.

The exploratory factor analyses were mostly in line with expectation. Raven, letter sets, and number series loaded onto a separate factor (Gf), as did rotation and symmetry span (WMC) and antisaccade, Stroop DL, flanker DL, and SACT (attention control). But one surprising result from the analysis was that the running letter span task loaded with the fluid intelligence measures and not the two complex span tasks. Previous studies have shown that running span correlates more strongly with Gf tasks than do complex span tasks, but to my knowledge studies from our lab have never shown running span to prefer Gf over WMC. This result could be because we did not include the customary operation span and instead had two spatial complex span tasks in symmetry and rotation span. The reason running span was not removed from the analyses altogether is because it was necessary for models to run properly. Removing it often resulted in one of two results – either only 1 or 2 factors would be extracted (instead of the hypothesized 3), or the model would fail to converge due to a communality exceeding 1. Therefore, one caveat with the following analyses is that running span is included despite it loading onto the theoretically wrong factor. Note that this specification problem was not present in the confirmatory factor analyses and structural equation models. In those analyses, model fit was good even when running letter span was omitted from the model.

In regard to the attention control factor, in contrast with Draheim et al. (2021b), I found that the Stroop DL task loaded much more strongly with the other attention tasks than did the flanker DL. Including flanker DL did not change the structure of the factors, and flanker DL had a clear preference to load with the other attention measures, but I opted to remove it from subsequent analysis because it generally failed to load meaningfully onto any factor.

Finally, in regard to visual arrays, each task loaded most strongly with the other attention control tasks when entered on their own (that is, without other visual arrays tasks; Figure 15). Overall, visual arrays tended to have stronger loadings to the other factors than most tasks as well, though. Further, the attention control loading for the visual arrays task tended to be smaller than the largest factor loading for most other tasks. Searching the literature for what constitutes an "acceptable" loading is challenging, because there are many rules of thumb and it strongly depends on the context. In my experience with cognitive tasks in studies like these, loadings around .33 or lower (< 10% variance attributable to a single factor) are unacceptably low, loadings from .33 - .40 (10-16% of variance) are poor, and loadings above .40 are at least adequate. We can see that nonselective visual arrays is just above the .33 line in terms of its loading onto attention control, suggesting it is a poor indicator of the construct in this dataset. The other visual arrays measures have loadings ranging from .41 - .49, indicating adequate-to-decent loadings in my opinion. A theme to the results is the relative strength of the concentric visual arrays measure, which was the task with the largest loading onto the "attention control" factor.

Next, I explored how the visual arrays behaved when entered along with other visual arrays tasks (Figure 16). I started by entering the color and concentric tasks with the 1:1 distractor-to-target ratio (VA4 and CVA; Figure 15a). This resulted in the first factor comprising the other three attention tasks and the two visual arrays measures, whereas in the previous analyses attention control had been relegated to the 2nd factor (behind Gf). Further, loadings were now quite strong for the visual arrays tasks to the attention control factor, and loadings were not strong with the WMC or Gf factors. I then entered all selective visual arrays tasks, that is I added the two enhanced versions (EVA4 and ECVA) to the previous model along with flanker DL (Figure 16b). Here, visual arrays began to dominate the first factor, as loadings for the antisaccade, SACT, and Stroop DL fell by about .10- .15 from what they were in previous models, and the selective visual arrays tasks had loadings ranging from .66 - .77. I next added the non-selective visual arrays (Figure 16c), which resulted in a factor comprised of just the five visual arrays tasks separating completely from the other attention control measures to occupy the first factor. For the final exploratory factor analysis, I removed most tasks and just entered the four selective visual arrays, the three Gf tasks, and the two complex span tasks (Figure 16d). Here, again visual arrays loaded strongly on the first factor, with Gf and WMC tasks on the other two factors.

The exploratory factor analysis results revealed that the constructs of attention control, WMC, and Gf are separable in these data, although this was obtained using analyses with oblique rotations in which the factors were fairly strongly correlated (generally around r = .40 - .45). The selective visual arrays variants generally behaved as attention control tasks, loading with the other attention control measures and not substantially with WMC. When the models were saturated with visual arrays tasks, the visual arrays measures began to separate from attention control, dominating the attention control factor when all selective versions were entered together (Figure 16b) and then completely separating from the other attention tasks when the non-selective visual arrays was entered (Figure 16c). This suggests that the bulk of the reliable variance in these visual arrays measures is shared with the other attention measures, but that there is some unique variance present in visual arrays performance as well. This could perhaps be an aspect of attention control not present in the other attention measures. I would offer that this could be a form of selective attention, however this would not explain why non-selective visual arrays loads so strongly with the other visual arrays tasks when entered together. So, this could be a method factor of little substantive value, *or* it could be an aspect of attention control that is common even to the non-selective visual arrays. But this is admittedly very speculative, and these results should not be over-interpreted.

a

	Factor 1	Factor 2	Factor 3
RAPM	.457	.116	-.197
LetterSets	.776	-.014	.062
NumberSeries	.735	-.057	-.125
RunLetter	.457	.096	-.023
SymSpan	.080	-.047	-.801
RotSpan	.018	.060	-.658
Antisaccade	.093	.756	.086
SACT	-.142	.490	-.133
StroopDL	.192	.474	.042
NSVAScore	.189	.339	-.230

Extraction Method: Principal Axis Factoring.
Rotation Method: Oblimin with Kaiser Normalization.

a. Rotation converged in 12 iterations.

b

	Factor 1	Factor 2	Factor 3
RAPM	.465	.086	-.193
LetterSets	.764	-.027	.039
NumberSeries	.740	-.052	-.118
RunLetter	.433	.156	.005
SymSpan	.057	-.069	-.884
RotSpan	.030	.106	-.591
Antisaccade	.114	.665	.059
SACT	-.163	.548	-.109
StroopDL	.156	.492	.038
VA4Score	.216	.460	-.180

Extraction Method: Principal Axis Factoring.
Rotation Method: Oblimin with Kaiser Normalization.

a. Rotation converged in 10 iterations.

c

	Factor 1	Factor 2	Factor 3
RAPM	.473	.096	-.194
LetterSets	.768	-.020	.051
NumberSeries	.743	-.071	-.122
RunLetter	.428	.139	-.004
SymSpan	.067	-.066	-.843
RotSpan	.016	.087	-.628
Antisaccade	.114	.686	.064
SACT	-.154	.532	-.119
StroopDL	.161	.472	.030
EVA4Score	.306	.408	-.215

Extraction Method: Principal Axis Factoring.
Rotation Method: Oblimin with Kaiser Normalization.

[illegible]

d

	Factor 1	Factor 2	Factor 3
RAPM	.480	.095	-.233
LetterSets	.774	-.003	.063
NumberSeries	.722	-.056	-.138
RunLetter	.417	.151	-.019
SymSpan	.101	-.033	-.738
RotSpan	-.016	.059	-.735
Antisaccade	.115	.658	.057
SACT	-.172	.557	-.113
StroopDL	.156	.498	.048
CVAScore	.226	.490	-.235

Extraction Method: Principal Axis Factoring.
Rotation Method: Oblimin with Kaiser Normalization.

a. Rotation converged in 11 iterations.

e

	Factor 1	Factor 2	Factor 3
RAPM	.533	.034	-.144
LetterSets	.776	.007	.067
NumberSeries	.731	-.038	-.115
RunLetter	.412	.143	-.020
SymSpan	.116	-.048	-.734
RotSpan	-.031	.084	-.726
Antisaccade	.130	.606	.047
SACT	-.164	.570	-.087
StroopDL	.147	.525	.037
ECVAScore	.178	.473	-.121

Extraction Method: Principal Axis Factoring.
Rotation Method: Oblimin with Kaiser Normalization.

a. Rotation converged in 9 iterations.

Figure 15 – Factor Structure with Each Visual Arrays Task Entered Separately. *Note*. Pattern matrices for exploratory factor analyses with each visual arrays task entered separately. (a) Non-selective visual arrays; (b) color visual arrays; (c) enhanced color visual arrays; (d) concentric visual arrays, (e) enhanced concentric visual arrays.

a

	Factor 1	Factor 2	Factor 3
RAPM	.154	.449	-.207
LetterSets	-.015	.776	.039
NumberSeries	-.002	.699	-.133
RunLetter	.133	.411	-.032
SymSpan	-.029	.087	-.758
RotSpan	.038	-.019	-.739
Antisaccade	.612	.095	.061
SACT	.555	-.189	-.075
StroopDL	.455	.141	.041
VA4Score	.681	.133	-.070
CVAScore	.712	.136	-.125

Extraction Method: Principal Axis Factoring.
Rotation Method: Oblimin with Kaiser Normalization.

a. Rotation converged in 7 iterations.

b

	Factor 1	Factor 2	Factor 3
RAPM	.139	.526	.130
LetterSets	-.017	.784	-.061
NumberSeries	.018	.714	.089
RunLetter	.086	.411	.057
SymSpan	-.036	.098	.750
RotSpan	.049	-.037	.752
Antisaccade	.494	.136	.006
SACT	.477	-.157	.127
StroopDL	.367	.158	.041
FlankerDL	.279	-.036	-.052
VA4Score	.730	.134	.063
CVAScore	.772	.140	.098
EVA4Score	.703	.221	.065
ECVAScore	.664	.109	.044

Extraction Method: Principal Axis Factoring.
Rotation Method: Oblimin with Kaiser Normalization.

a. Rotation converged in 6 iterations.

c

	Factor 1	Factor 2	Factor 3	Factor 4
RAPM	.362	.463	.092	-.135
LetterSets	.010	.761	-.050	.072
NumberSeries	.219	.645	.073	-.113
RunLetter	-.068	.449	.109	.190
SymSpan	.122	.120	.554	-.043
RotSpan	-.078	-.086	1.010	.003
Antisaccade	.212	.150	.045	.434
SACT	.217	-.127	.136	.334
StroopDL	-.016	.248	.050	.568
FlankerDL	.151	-.039	-.025	.166
VA4Score	.824	.032	.028	.062
CVAScore	.704	.080	.113	.150
EVA4Score	.761	.132	.055	.061
ECVAScore	.570	.055	.078	.174
NSVAScore	.837	.016	.025	-.096

Extraction Method: Principal Axis Factoring.
Rotation Method: Oblimin with Kaiser Normalization.

a. Rotation converged in 8 iterations.

d

	Factor 1	Factor 2	Factor 3
RAPM	.156	.564	.044
LetterSets	-.034	.739	-.033
NumberSeries	-.001	.761	.055
SymSpan	-.084	.081	.864
RotSpan	.101	-.047	.628
VA	.805	.016	.062
EVA	.820	.117	-.021
CVA	.919	-.057	.070
ECVA	.778	.006	-.050

Extraction Method: Principal Axis Factoring.
Rotation Method: Oblimin with Kaiser Normalization.

a. Rotation converged in 5 iterations.

Figure 16 – Factor Structure with Visual Arrays Entered Together. *Note.* NSVA = non-selective visual arrays; VA4 or VA = color selective visual arrays; EVA4 or EVA = enhanced color selective visual arrays; CVA = concentric visual arrays; ECVA = enhanced color selective visual arrays; SACT = sustained attention-to-cue; RAPM = Raven's advanced progressive matrices; RotSpan = rotation span; SymSpan = symmetry span.

2.2.6 Predictive Validity and Relationship Among Factors

The next set of analyses were conducted to assess the relationship among the factors as well as the predictive validity of the visual arrays tasks. Interestingly, in confirmatory factor analysis I did not have the specification problem earlier noted with regard to the letter running span task. Instead, model fit was just fine when letter running span was omitted from the models, allowing robust factors of WMC (symmetry and rotation span) and Gf (Raven, number series, and letter sets) to emerge, in addition to separate factors for attention control (antisaccade, SACT, and Stroop DL) and visual arrays (all tasks with or without the non-selective version). The visual arrays factor was strongly correlated with, but separable from, the factor consisting of the other attention measures (Figure 17; also see Figure 16c).

The first confirmatory factor analysis was conducted to assess the relationships among the factors (Figure 17). Attention control, visual arrays (selection tasks only), WMC, and Gf had relatively the same relationship to one-another, around r = .60 or 36% shared variance at the latent level. The exception is with the selective visual arrays and the attention control factor, which shared 2/3 of its variance. However, the path between the two was statistically different from 1 and model fit was good, indicating that VA-S and attention control are separable factors. Rephrased, this again indicates that a component of visual arrays performance is unique to the paradigm.

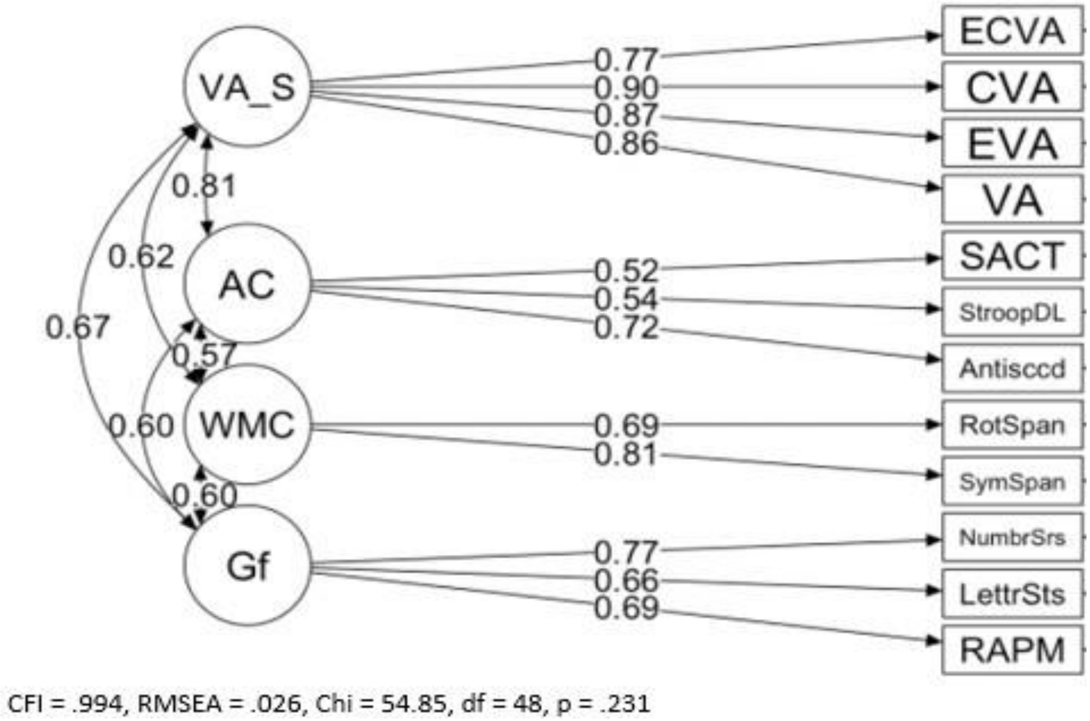

Figure 17 – Confirmatory Factor Analysis Showing Relationship Among Factors. *Note.* VA_S = selective visual arrays; WMC = working memory capacity; Gf = fluid intelligence; ECVA = enhanced concentric visual arrays; CVA = concentric visual arrays; EVA = enhanced color selective visual arrays; VA = color selective visual arrays; SACT = sustained attention-to-cue; RotSpan = rotation span; SymSpan = symmetry span; RAPM = Raven's advanced progressive matrices.

The next model tested whether attention control fully mediated the WMC-Gf relationship. A full mediation is predicted under the executive attention view of individual differences in WMC (e.g., Engle, 2002, also refer to the introduction), but generally does not emerge in large-scale correlational studies. We were able to find a full mediation in Draheim et al. (2021b), which was shown to be only achievable when attention control was measured using accuracy-based measures, as they are in the present study. The first analysis I conducted was a stepwise regression WMC, attention control, and selective visual arrays factor scores predicting Gf factor scores. The results showed that these factor scores accounted for 42% variance in the Gf factor scores, but that all three had statistically significant independent contributions Gf (7% for VA-S, and 2.2% each for attention control and WMC). I also conducted a more direct test using structural equation modeling and a full mediation did not occur in the present dataset. Around 15% of the variance in the WMC-Gf relationship was unaccounted for by attention control, which was statistically different from 0 in this good-fitting model (Figure 18). Note that the model shown in Figure 18 shows the three non-visual arrays tasks as indicators of attention control, but models including visual arrays tasks (not shown) also did not achieve full mediation. It is not immediately clear why a full mediation was found in our previous study but not this one, however one difference in the studies is that operation span was not included in the present dataset. We instead opted for verbal running span tasks, one of which did not load with the complex span and the other was in session 4 and so not analyzed. This perhaps changed the nature of the WMC factor to involve more spatial processing than when verbal tasks are included, and perhaps this spatial processing was unique to the WMC and Gf tasks.

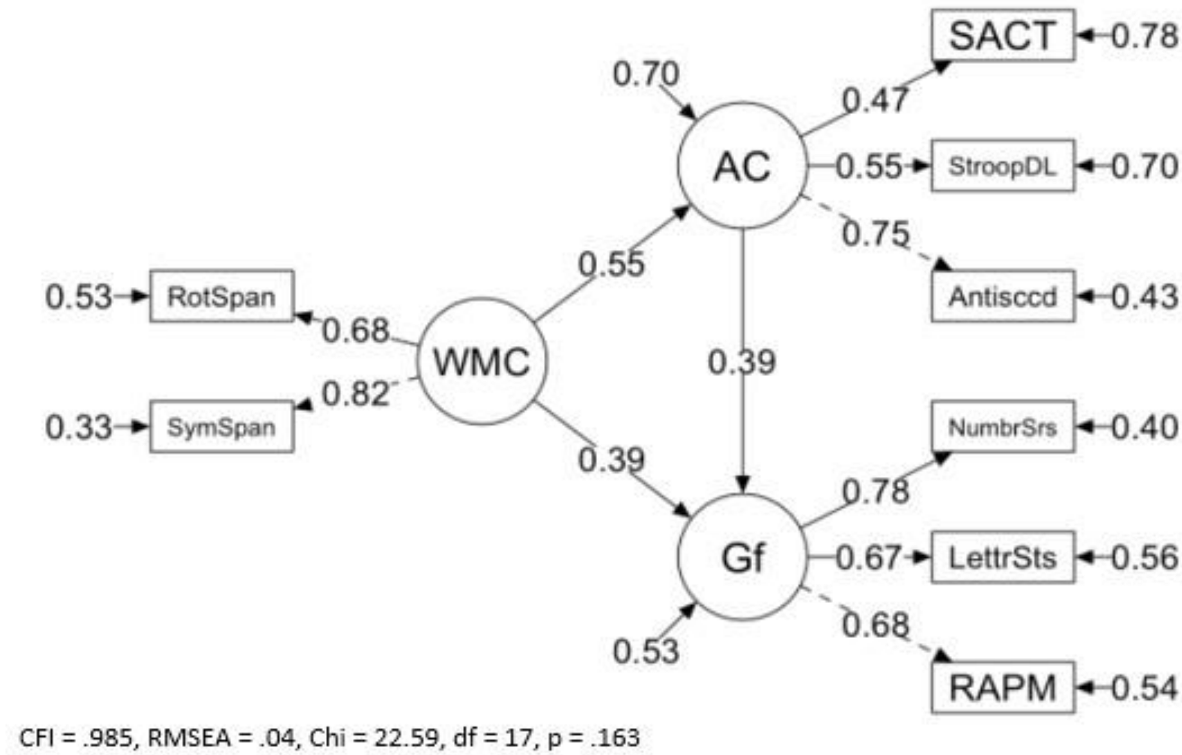

Figure 18 – Attention Control Mediating the WMC-Gf Relationship. *Note.* WMC = working memory capacity; AC = attention control; Gf = fluid intelligence; SACT = sustained attention-to-cue; Antisccd = antisaccade; RotSpan = rotation span; SymSpan = symmetry span; RAPM = Raven's advanced progressive matrices.

Next, I assessed the relative contribution of attention control and visual arrays to WMC and Gf (Figure 19). In the model, all visual arrays tasks load onto a separate factor from the other attention control measures, and both factors predict WMC and Gf. Notably, the visual arrays and attention control factors are allowed to correlate, which they do substantially at $r = .79$ (62% shared variance). However, the visual arrays factor predicted 19% incremental variance in WMC and 28% incremental variance in Gf above and beyond attention control. On the other hand, attention control did not predict any incremental variance in these constructs above and beyond the visual arrays tasks (the paths of .21 to

WMC and .18 to Gf are not statistically different from 0). This indicates that there is a unique aspect of performance in the visual arrays *and* that this variance is meaningful, as it is shared with WMC and Gf tasks. Given that non-selective visual arrays is in this model and has a decent loading (.75), this unique aspect of visual arrays appears to be more so related to memory storage than selective attention.

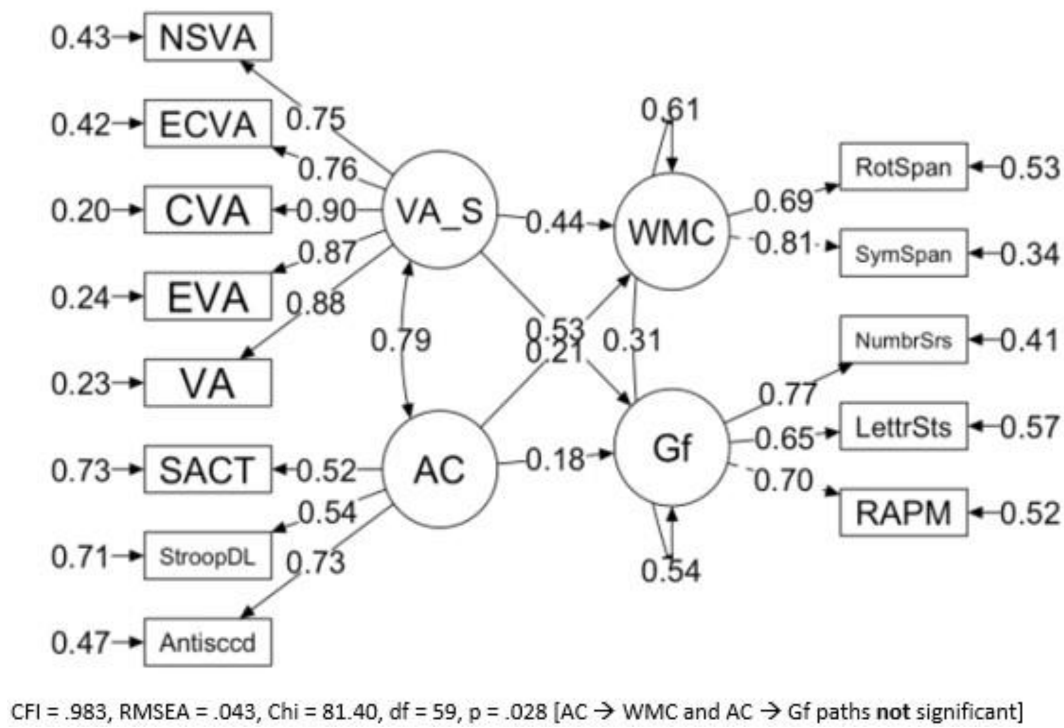

Figure 19 – Attention Control and Visual Arrays Predicting WMC and Gf. *Note.* VA_S = visual arrays including non-selective; AC = attention control; WMC = working memory capacity; Gf = fluid intelligence; NSVA = non-selective visual arrays; ECVA = enhanced concentric visual arrays; CVA = concentric visual arrays; EVA = enhanced color selective visual arrays; VA = color selective visual arrays; SACT = sustained attention-to-cue; Antisccd = antisaccade; RotSpan = rotation span; SymSpan = symmetry span; NumbrSrs = number series; LettrSts = letter sets; RAPM = Raven's advanced progressive matrices.

Informed by the results of the exploratory factor analysis, I was able to calculate estimated factor scores using the regression method. Factor scores are akin to composite z-scores, except that tasks are differentially weighted when using factor scores (see DiStefano et al., 2009). That is, factor scores estimate an individual's relative position on the overall *factor*, whereas z-score composites indicate each individual's z-score averaged across each task, which are weighted equally. The benefit to the factor score approach is that it is multivariate and so scores on tasks which load more strongly with the factor (e.g., antisaccade compared to Stroop DL) contribute more to the score than they would with a composite z-score, providing a more precise estimate of ability. One disadvantage is that factor scores can only be calculated for participants who have valid data points for every task included in the factor analysis used to obtain them (not just every task of that construct, as with the standard z-score composites), and so data are lost. To this point, I was able to calculate factor scores for just 168 participants on WMC (rotation and symmetry span), Gf (Raven, letter sets, number series), and attention control (antisaccade, SACT, Stroop DL) and 189 for the selective visual arrays factor (all visual arrays tasks minus the non-selective version).

For criterion validity, I used both Gf and multitasking. Note that multitasking is likely a better criterion in theory (e.g., Martin et al., 2020), as they are designed to be a proxy for real-world performance. However, multitasking measures may not exhibit the best psychometric properties, in part because scoring is arbitrary and often not clear to the participants. Further, in this dataset, there were only two multitasks and it was not possible to obtain factor scores for them – they would not load onto a separate factor from either the Gf or attention control tasks, depending on the model. As such, I will show results predicting both Gf and multitasking. WMC was also added into the models for completion, although it is worth nothing that WMC is not a typical criterion measure. Also note that correlations between each visual arrays task and each potential outcome or discriminant validity measure is presented in Appendix A.

Semipartial correlations from regression analyses are shown on Table 3. The picture is clear – the selective visual arrays factor score (VA-S) and attention control factors both predict incremental variance in Gf, WMC, and multitasking. The trend was for the VA-S factor scores to predict more unique variance than attention control, except that attention control had a numerically (but not statistically) stronger semipartial correlation to WMC than VA-S. This is quite interesting given that visual arrays is generally considered a WMC task whereas the attention control measures (antisaccade, SACT, and Stroop DL) have virtually no storage or memory component/demands to them. Also note this is counter to the previous structural equation modeling showing that attention control did not predict any unique variance in Gf or WMC above and beyond the visual arrays factor. This could be because analyses using path modeling and analyzes involving correlating factor scores will not produce the exact same results. It could also be because non-selective visual arrays was included in the structural equation model but not the factor score calculation.

Table 3 – Semipartial Correlations between Attention Control and Visual Arrays Regressed on Various Criteria.

Construct	AC	VA-S	Total R
Fluid Intelligence	.20*	.32*	.64*
Working Memory Capacity	.26*	.24*	.61*
Multitasking	.22*	.26*	.59*

Note. AC = attention control factor scores; VA-S = selective visual arrays factor scores. Fluid intelligence and working memory capacity are factor scores. Multitasking is the composite z-score of synthetic work and the Foster task. The AC and VA-S factor scores correlate at $r = .66$. *$p < .05$

Next, I assessed to what extent visual arrays performance predicts variance in outcome measures above and beyond processing speed. The motivation for these analyses is a common concern that the reason accuracy-based attention control measures correlate strongly with other cognitive measures is because of factors or processing that are not directly attributable to attention control, such as general processing efficiency, motivation, general cognitive ability, etc. Analyzing the semipartial correlations shown on Table 4, it is clear that both processing speed and VA-S have substantial unique contributions to both Gf and multitasking. Further, VA-S shares 22% incremental variance in WMC over processing speed, whereas processing speed does not have a statistically significant unique prediction to WMC (the semipartial correlation of r = .17 is non-significant due to low sample size for the processing speed tasks). Although not shown, what is interesting is that if you add the attention control factor scores into the model, it does not predict any additional variance in any of these criterion measures.

I do not want to overinterpret these results, as there were only 97 participants who had scores on all processing speed measures (they were in session 4 of the study at large). However, so far it appears that while our processing speed measures are strongly correlated with both Gf and multitasking, visual arrays performance (that is, attention control) accounts for a substantial amount of variance above and beyond processing speed. That is, the attention control tasks are not "just" processing speed or general ability. When the data are fully collected, we plan to investigate further the reason for processing speed's such large correlation with, in particular, Gf. One hypothesis is that processing speed tasks, particularly the ones we administered, are functionally attention control measures (see Mashburn et al., 2020).

Table 4 – Semipartial Correlations between Processing Speed and Visual Arrays with Various Criteria.

Construct	PS	VA-S	Total R
Fluid Intelligence	.30*	.44*	.74*
Working Memory Capacity	.17	.47*	.65*
Multitasking	.45*	.33*	.74*

Note. PS = processing speed factor scores; VA-S = selective visual arrays factor scores. Fluid intelligence and working memory capacity are factor scores. Multitasking is the composite Z-score of synthetic work and the Foster task. The PS and VA-S factor scores correlate at $r = .46$. *$p < .05$

2.2.7 Differences in Performance Across Set Sizes

One of the primary questions of this study was the nature of individual differences in performance across different set sizes. This question was motivated in part by results from Fukuda, Vogel and colleagues. Fukuda et al. (2015) hypothesized that as the number of items to be maintained in primary memory exceeds one's capacity, additional selective attention resources are required to properly manage and maintain the number of items equal to that individual's capacity. Failure to do so would result in possibly a catastrophic drop in overall performance which they referred to as an overload of task-relevant information. Fukuda et al. found that span differences (differences in performance between low- and high-WMC individuals) in a location-based (hemifield) selection visual arrays task were nonexistent for smaller set sizes but then grew as set sizes increased. Specifically, both k scores and contralateral delay activity (indicator of how much information is being maintained in primary memory) increased with set size for high-WMC individuals until around 3-4 items and then stabilized, not decreasing except a little in the largest set size (8). Conversely, k scores and contralateral delay activity in low-WMC individuals peaked at set size 3 and then declined with each increasing set size. Differences between high- and low-WMC individuals in *k* scores and contralateral delay activity were minimal and possibly not statistically significant at set size 3, but were large and obviously significant for set size 5.[4] This informed my decision to use set size 3 and 5 in the present study, specifically, any set size lower than 3 would likely result in ceiling effects and minimal individual differences in these tasks, anything above 5 or 6 would clutter the screen in the enhanced versions of the task, particularly the concentric version, and perhaps be overly difficult (to that end, the lowest accuracy within any trial type in the present study was object-based allocation trials in set size 5 of the ECVA task, which had a 67.5% average accuracy rate, only 17.5% above chance performance). As such, set sizes 3 and 5 seemed to hit the sweet spot such that neither were too easy or too difficult, thus providing good differentiation between high- and low-ability individuals in service of maximizing predictive validity and potentially providing insights as to how attention control is applied in the visual arrays paradigm.

Unfortunately, this does not appear to be the case in the present data. I created factor score estimates for set size 3 and set size 5 performance in the selective visual arrays tasks using the same methods I used to obtain factor scores for some of the other constructs and then tested the extent to which they predict the other constructs (Table 5). First, performance across set size 3 and set size 5 was very strong, at r = .87 or 76% shared variance among the factor scores. I also attempted to run structural equation models and set

[4] Fukuda et al. (2015) did not report significance testing for their results, but they did show a graph with 95% confidence intervals (see their Figure 4b and 4c). Visual inspection seems to indicate that span differences were significant, though relatively small, at set size 3 for *k* scores and possibly not statistically significant for contralateral delay activity.

size 3 and 5 performance was not possible to separate, for instance one model I tried would not run fully and showed a path estimate of .97 between the set size 3 and set size 5 factors. Further, correlations between set size 3 and set size 5 factor scores with the various outcome measures were very similar and not statistically different one another. The exception was that set size 5 performance had a statistically stronger relationship to WMC than did set size 3 performance, which is not surprising given that more information must be maintained in primary memory for set size 5 as opposed to 3. I would have expected that difference to be larger, and for set size 5 to generally be more predictive than set size 3. Note that *k* scores were predictable larger for set size 5 than set size 3 (by .28 of a *k* score, Cohen's d = 0.57, $p < .001$) but that accuracy rates were much lower for set size 5 (74%) than set size 3 (85%; Cohen's d = 2.49, $p < .001$). This means that that set size 3 and set size 5 performance had massive differences in difficulty but no meaningful differences in their relationship to other cognitive variables. I had planned to analyzed interactions with set size and other manipulations, but this seemed moot after these results. Though I did check to see whether this result could possibly be because the enhanced versions of the tasks were perhaps too demanding on participants, and this did not seem to be the case as even in the easier tasks (VA4 and CVA), set size 5 performance was not more predictive than set size 3 (if anything, set size 3 appeared more predictive).

Table 5 – Comparing Set Size 3 and Set Size 5 Performance to Various Criteria.

Construct	k3	k5	Total R
Attention Control	.23*	.15	.67*
Fluid Intelligence	.08	.29*	.63*
Working Memory Capacity	.01	.27*	.58*
Multitasking	.16*	.12	.57*

Note. k3 = factor scores from set size 3 in all selective visual arrays tasks; k5 = factor scores in from set size 5; *p* value from the two-tailed test of whether the two correlations are statistically significantly different using the Williams T2 statistic (Steiger, 1980). Attention control, fluid intelligence and working memory capacity are factor scores. Multitasking is a composite Z-score of synthetic work and the Foster task. The k3 and k5 scores correlate at $r = .87$. $*p < .05$

This result was discouraging but possibly highly informative. Given the large differences in accuracy between the different set sizes, and that set size 5 performance was under halfway between chance and perfect performance, I can reasonably rule out that these results are because set size 5 trials were not challenging or demanding enough. Instead, it appears set size 3 performance is more demanding than expected. So, the take-home message is that even with the relatively small set size of 3 that is hypothetically near and below most individual's true capacity, there are strong individual differences in the selective visual arrays tasks – and these individual differences correlate very strongly with other attention control measures as well as measures of other abilities.

After rereading Fukuda et al. (2015), I realized there were critical differences between their methodology and mine that could explain the discrepancy in results. The simplest is that they only had 36 participants (but 1600 trials per participant!) in their second experiment, which is the one in which they showed that performance was relatively similar between high- and low-spans for set size 3 but that there were large differences in set size 5 performance. Beyond that, however, Fukuda et al. appears to have used an easier version of visual arrays than the variants used here. First, their task was selective based on hemifields of the screen, which has been shown to be easier than feature-based selection (e.g., Anllo-Vento & Hillyard, 1996; Vogel et al., 2015). Second, the change to be detected by their participants was a change in color and not a change in orientation as it was here, which may be easier as it would presumably take less processing to determine a change. Because their task might have been easier in this regard, this could explain why they needed larger set sizes to show span differences whereas in the present study set size 3 was sufficiently difficult for most participants and did not exacerbate individual differences. However, another key difference is that Fukuda et al. presented the target (initial) array on the screen for just 150 ms. In the present study, the target array was presented for 250 ms in the color selective versions and 300 ms for the concentric versions. I realized in hindsight that the difference in durations could complicate comparisons between these two tasks (thankfully it does not appear to have had too negative of an effect, as accuracy was lower for the concentric versions plus the CVA task was the most predictive). More to the point, the target arrays were shown to the participants for much longer in my design than Fukuda et al. This could have resulted in some manipulations not having the intended effect. For example, in Experiment 3 of Fukuda et al. they manipulated how long participants had to view the target array, either 150 ms, 300 ms, or 450 ms and found that high-WMC individuals showed essentially no improvement in performance as presentation time increased, but low-WMC individuals did – specifically improving the most from 150 ms to 300 ms in the largest set size (8). In the present design, it is not clear whether 250/300 ms array exposures were too long such that individual differences with increasing set sizes were attenuated, but it is possible and consistent with Fukuda et al.'s findings.

2.2.8 Effect of Distractors

I next assessed to what extent the presence and number of distractors contributed to attention-relevant individual differences in the visual arrays tasks. The easiest way to test this was to look at task-level differences in the non-selective visual arrays, color selective visual arrays, and enhanced color selective visual arrays – as these tasks differ only in the number of distractors present. Specifically, non-selective has no distractors, color selective has a 1:1 ratio between distractors and targets and the enhanced version has a 2:1 ratio of distractors to targets.

Performance was expectedly worse as the number of distractors increased (see Table 1 for mean scores from the tasks). The difference between NSVA and VA4 had an effect size of .67, and the difference between VA4 to EVA4 had an effect size of 0.34 (Cohen's d), and these differences were statistically significant. It is interesting that going from 0 distractors to a 1:1 distractor to target ratio had about twice the impact on difficult as going from a 1:1 to 2:1 distractor to target ratio.

Stepwise regression models showing the correlation to other measures as distractors increase are shown on Table 6. In every case, adding distractors resulted in a statistically stronger relationship to the criterion. These effects were strongest for attention control and multitasking, specifically adding distractors resulted in 10% more shared variance with attention control and 11% more variance in multitasking. Going from a 1:1 distractor to target ratio to a 2:1 ratio generally resulted in slightly less shared variance in the outcome variables, for example 6% incremental variance in predicting attention control and 8% in predicting multitasking.

These results are not surprising nor novel, as a number of studies have shown that adding distractors resulted in performance reflecting a greater degree of attention-relevant individual differences (e.g., Fukuda and Vogel, 2009; Martin et al., in press; Vogel et al., 2005). What is interesting is that overall, the improvement in predictive validity between visual arrays and both Gf and multitasking from doubling the number of distractors was around the same magnitude as the improvement in predictive validity in going from no distractors to some. This may indicate a linear trend such that each distractor added will roughly improve prediction to the same degree, likely up to a certain point. However, this is speculative, and the present study was not designed to adequately assess this question.

Table 6 – Incremental Criterion Validity by Adding Distractors.

Construct	No Distractor	1:1 Distractor Added	2:1 Distractor Added
Attention Control	24%*	10%*	6%*
Fluid Intelligence	19%*	6%*	8%*
Working Memory Capacity	19%*	6%*	5%*
Multitasking	12%*	11%*	8%*

Note. AC = attention control factor scores; Gf = fluid intelligence factor scores; WMC = working memory capacity factor scores; MTz = multitasking z-score composite. No Distractor is shared variance solely with the non-selective visual arrays task; 1:1 Distractor Added is the incremental variance when the color selective visual arrays is added to the regression model; and 2:1 Distractor Added is the incremental variance when the enhanced color selective visual arrays is added to the regression model. *$p < .05$

2.2.9 Feature vs. Location Distraction

The next set of analyses were designed to answer the question as to whether feature-based (color) selection and location-based are qualitatively and quantitatively different forms of selection.

First, the two location-selection visual arrays tasks were markedly more difficult than the non-selective visual arrays task (mean difference of *k* scores was 0.37 between CVA and NSVA, Cohen's d = 0.59; mean difference of *k* scores was 0.84 between ECVA and NSVA, Cohen's d = 1.27), and performance on location-selection trials correlated substantially with other cognitive measures above and beyond non-selective visual arrays (roughly 20% incremental variance to attention control factor scores and 12% to fluid intelligence). This indicates that making a location-based selection is more difficult and effortful, and that there are strong individual differences in the ability to make this type of selection.

Connecting to the broader literature, several studies have reported that selection via features (e.g., color) is more difficult than selection via location within the visual arrays paradigm (e.g., Anllo-Vento & Hillyard, 1996; Vogel et al., 2015). In the present study, this was not the case. The CVA task was numerically more difficult than its counterpart (color selective visual arrays), albeit this difference was negligible (Cohen's d = 0.06) and failed to reach statistical significance in a paired samples t-test (two-tailed $p = .40$). On the other hand, mean scores were quite different between the enhanced color visual arrays and the enhanced concentric visual arrays (Cohen's d = 0.54; two-tailed $p < .001$). It is worth noting again that studies showing that feature-based selection is more difficult than location-based tend to use location selection via sides or corners of the screen whereas spatial selection in the present study was designed to be intentionally more demanding. Furthermore, although participants received extensive practice on the concentric visual arrays tasks, and care was taken to minimize the extent to which it was ambiguous which items were in the target location vs. the to-be-ignored location, it cannot fully be ruled out that participants in the present study were not always certain whether a target was in the to-be-attended region or the to-be-ignored region. Any uncertainty from the participant could add noise to the data and result in lower k scores.

Next, we can ask whether feature-based selection and location-based selection are qualitatively different in nature. That is, does performance on trials with these types of demands have different relationships to other variables? The most straightforward way to address this is to compare scores on the base versions of the visual arrays tasks which require feature-based selection (VA4) and location-based selection (CVA). I conducted several regression models showing that the answer to this is an enthusiastic yes – even though the VA4 and CVA tasks correlated strongly and load onto the same factor (see the exploratory factor analyses), the CVA task shares a substantial amount of *incremental* variance with the attention control factor scores (8.7%), the Gf factor scores (5.9%), the WMC factor scores (6.0%), and, to a lesser extent, the multitasking z-score composite (3.1%) over and above the VA4 task. On the other hand, VA4 did not statistically share any incremental variance in Gf or WMC above and beyond the CVA task, and only a little bit in attention control and multitasking (Table 7 shows the semipartial correlations, note that due to rounding the values in the table will not perfectly match the % variance reported here in the text). In other words, the baseline tasks for feature-based selection (VA4) and location-based selection (CVA) correlated strongly (r = .80), but CVA contributed substantial unique variance to several criterion measures above and beyond VA4, whereas the reverse was not true. This is more impressive given that CVA was the third visual arrays task participants performed, and so they were generally well-practiced in the paradigm.

Table 7 – Semipartial Correlations between VA4 and CVA on Several Criterion Measures.

Construct	VA4	CVA	Total R
Attention Control	.13*	.30*	.64*
Fluid Intelligence	.13	.24*	.55*
Working Memory Capacity	.12	.24*	.54*
Multitasking	.16*	.18*	.52*

Note. VA4 (color selective visual arrays) and CVA (concentric visual arrays) regressed onto various outcome measures. Attention control, fluid intelligence, and working memory capacity are factor scores. Multitasking is the composite z-score of synthetic work and the Foster task. CVA and VA4 correlated at r = .80. *p < .05

However, I also tested whether the results were similar when comparing the enhanced versions of the visual arrays task. Table 8 shows that, the pattern of results is essentially flipped when analyzing just the enhanced tasks. That is, the EVA4 task predicted roughly 10% incremental variance in attention control, Gf, WMC, and multitasking above and beyond the ECVA task, whereas the ECVA task only had one statistically significant semipartial correlation to any of these criteria (r = .19 with attention control). Further, when I created factor scores for color selection (using trials from the VA4 and EVA4 tasks) and location selection (using trials from the CVA and ECVA tasks), the results were similar in that color selection had a statistically significant semipartial correlation to all four criteria whereas location selection only had a statistically significant semipartial correlation to attention control (Table 9). It should be noted that the factor scores for location selection appeared to weight ECVA performance more strongly, as factor loadings (in the pattern matrix) were in the high .80s for those trials compared to the mid .60s for the CVA task. These results may therefore be due to the relatively poor criterion validity displayed by the ECVA task overall, perhaps because the task was overly difficult.

Table 8 – Semipartial Correlations between EVA4 and ECVA on Several Criterion Measures.

Construct	EVA4	ECVA	Total R
Attention Control	.32*	.19*	.62*
Fluid Intelligence	.35*	.11	.57*
Working Memory Capacity	.34*	.07	.51*
Multitasking	.31*	.12	.55*

Note. EVA4 and ECVA tasks correlated $r = .68$. $*p < .05$

Table 9 – Semipartial Correlations for Factor Scores of Color- and Location-Based Selection.

Construct	Color	Location	Total R
Attention Control	.22*	.18*	.65*
Fluid Intelligence	.24*	.10	.55*
Working Memory Capacity	.27*	.04	.51*
Multitasking	.23*	.08	.52*

Note. Color = factor scores for performance in the VA4 and EVA4 tasks; Location = factor scores for performance in the CVA and ECVA tasks. Color and spatial factor scores correlated $r = .83$. $*p < .05$

Anllo-Vento & Hillyard (1996) argued that selection based on features (color) and selection based on location were qualitatively different from one another and involved separate attentional mechanisms, which they argued was consistent with ideas expressed by Treisman and colleagues regarding the feature integration theory (e.g., Treisman, 1993). Their claim was based on findings that behavioral spatial selection and color selection in visual arrays was associated with different event-related potentials in their electroencephalogram experiments with an N of 12. Specifically, selection via location evoked early positive and negative (P1 and N1) components beginning around 80 ms and 140 ms, respectively, after presentation, whereas selection via color evoked broad selection negativity and selection positivity components (SN and SP) that had later onsets and persisted longer. In the present study, there was some evidence to suggest that color selection and spatial selection are qualitatively different. Relative to the feature-based selection task (VA4), the concentric visual arrays shared 9% unique variance with the attention control factor score (antisaccade, SACT, and Stroop DL), predicted 6% incremental variance in the Gf factor score, and predicted 3% incremental variance in the multitasking composite score. These findings of both the present study and Anllo-Vento and Hillyard are perhaps consistent with the theoretical framework I am operating with that attention control is domain-general, but may be applied differently in different situations. On the other hand, the results were essentially flipped when comparing the two enhanced versions of the tasks – and analyses using factor scores for location- and color-based selection suggested that color-based selection trials were roughly 5% more predictive on average to the various criterion measures. It should also be noted that the color-selective and location-selective tasks all loaded onto the same factor in exploratory factor analysis, and their respective factor scores correlated very strongly at r = .83.

2.2.10 Differences in Spatial Configuration of Stimuli

Another primary question of the present study was the extent to which individuals differ in their ability to allocate visual attention to stimuli which differ in their spatial configuration. These analyses were done using the two concentric visual arrays tasks (CVA and ECVA). Reminder that these tasks have qualitatively different types of stimuli configurations, and that these configurations were more controlled than in the color-selective VA tasks (in which stimuli can appear in any location on the screen provided they did not touch or overlap other stimuli). In the CVA task, targets could appear either within 3° of the center of the screen, with distractors on the periphery (on a ring 9.3° from the center), or the reverse. In the ECVA task, targets could appear focally, on a ring 9.3° from the center, or on a ring 18.5° from the center, with distractors occupying the regions in which the targets did not appear. Trials in which targets were at the center and distractors were peripheral will be labeled *spotlight* trials. Trials in which the distractors are at the center and targets peripheral will be labeled *donut* trials, as I hypothesized that they would require different types of visual allocation (spotlight vs. donut). Finally, trials in the ECVA in which targets are on the middle (9.3°) ring with distractors both inside and outside of this ring will be labeled as object-based, as these trials may be qualitatively different from the other donut trials since distractors appear both inside and outside of the target region.

Expectedly, analyses show that donut trials were substantially more difficult than spotlight trials. In the CVA task, the mean difference in k scores was .36 (Cohen's d =

0.56), and in the ECVA task the difference was .76 (Cohen's d = 0.94). To put these results into context, the mean difference in performance between donut trials and spotlight trials was about twice as much as the mean difference between spotlight trials and performance in the non-selective visual arrays task. In other words, donut allocation was twice as difficult relative to spotlight trials as spotlight allocation was to no selection at all. In the ECVA task, there were no statistically significant differences in the relative difficulty of donut trials as opposed to object-based trials. That is, the two types of non-spotlight configurations in ECVA were equally difficult, statistically speaking.

I calculated factor scores for spotlight trial performance and donut trial performance from both concentric visual arrays tasks (Table 10). Interestingly, both spotlight and donut trials had statistically significant semipartial correlations to attention control, Gf, WMC, and multitasking, except in one instance in which spotlight trials did not have a significant semipartial correlation to WMC. Donut trial performance has a numerically stronger semipartial correlation to each criterion than did spotlight trial performance, and this was largest for the relationship to WMC, of which donut trial performance shared an additional 14% (!) variance above and beyond spotlight trial performance. The combined findings that donut trial performance was much lower and more predictive than spotlight trial performance support the hypothesis that donut allocation is a more effortful and difficult type of allocation. However, it is worth mention that performance in the object-based trials in the ECVA task (i.e., trials with targets along the 9.3° ring and distractors are both in the center of the screen and outside of the targets) generally correlated the weakest to attention control, Gf, WMC, and multitasking than any other trial type among all the visual arrays tasks, including the non-selective version. As a result, when analysis was restricted only to the ECVA task, spotlight trial performance was overall more predictive than object-based performance. I had hoped in designing the ECVA task that performance on the object-based allocation trials would prove to be challenging and show meaningful individual variation. It turns out that only the former was correct. It is possible that this is a floor effect, as performance for the middle trials was just 67.5% in set size 5 (77.4% in set size 3, note that 50% is chance performance in these tasks). Discounting the results from the ECVA task on object-based trials, these results support the argument from Bleckley et al. (2003; 2014) that individuals of higher cognitive ability are more efficient and flexible with their visual allocation.

Table 10 – Semipartial Correlations for Spotlight and Donut Trials.

Construct	Spot	Donut	Total R
Attention Control	.20*	.22*	.66*
Fluid Intelligence	.14*	.24*	.60*
Working Memory Capacity	.04	.38*	.58*
Multitasking	.13*	.25*	.58*

Note. Spot = spotlight trial factor scores from CVA and ECVA. Donut = donut trial factor scores from CVA and ECVA. Spotlight and donut scores correlate $r = .79$. $*p < .05$.

2.2.11 Differences in Performance Across Preparation Time

The final question of interest is to what extent individual differences emerge, and interact with, participants having more or less time to prepare for the initial array. Recall that in the concentric VA tasks, the cue appeared for 250 ms and was followed by a variable cue-to-stimulus interval that was either 100 ms, 400 ms, or 700 ms. As such, participants on some trials had as little as 350 ms and as much as 950 ms from cue onset to onset of initial array. Note that each cue-to-stimulus interval appeared randomly but with equal weighting across all trial types. Also, participants were not aware of what the cue-to-stimulus would be on any given trial, but they were informed that the delay between the cue and array would be shorter in some trials and longer in others, and were encouraged to use the extra time to prepare for the upcoming array.

Differences in *k* scores as a function of cue time are shown on Table 11. Allowing additional time for participants to prepare for the target array had only a minor effect on performance. Specifically, the *k* difference between a 100 ms and 700 ms cue-to-stimulus interval was only .15 (Cohen's d = 0.23, p <.001) for the concentric task and was not statistically significantly different in the enhanced concentric task. Further, performance in the enhanced task was surprisingly best in the 400 ms condition.

I created factor scores for the different cue-to-stimulus intervals and then conducted regression analyses to see whether less preparatory time in the CVA and ECVA tasks resulted in a stronger relationship to attention control, Gf, WMC, and multitasking (Table 12). Generally, the answer is no. In fact, semipartial correlations were strongest overall strongest for the 700 ms cue-to-stimulus interval trials. Therefore, my prediction that increasing the cue-to-stimulus interval would reduce individual differences was not supported in this dataset.

This was a surprising finding and is inconsistent with Heitz and Engle (2007) who found that high- and low-WMC individuals differed in the amount of time it took them to reach asymptotic accuracy in an Eriksen flanker task. They found the largest span differences around the 400 – 600 ms window, meaning that when RTs were forced to be in that range (using a response deadline), there were large differences in accuracy between high- and low-WMC participants. They concluded that this finding was because individuals of differing cognitive ability require longer times to constraint their attention (in a spotlight manner) down to attend to just the focal stimuli and not the peripheral distractors. In the present study, even though the cue-to-stimulus interval was 100 – 700 ms, participants had a total of 650 – 1250 ms to prepare for and process the array. To explain, in the shortest cue-to-stimulus trials, the cue was presented for 250 ms, the cue-to-stimulus interval was 100 ms, and the target array was presented for 300 ms. Perhaps this was sufficient for most individuals to prepare for the trial and constrain their attention accordingly and is why performance in the shortest cue-to-stimulus trials were not generally more predictive than performance on trials with longer intervals.

Another possibility is that participants did not need much preparatory time given the fairly long (300 ms) presentation time of the target array in the concentric tasks. That is, perhaps the target array was displayed for a sufficiently long period such that participants did not feel the need to use the extra time to prepare afforded in the 400 ms and 700 ms cue-to-stimulus intervals even though they were encouraged to do so via instruction.

Differences between Heitz and Engle's (2017) paradigm and the present design cannot be ruled out as a possibility for discrepant results. Heitz and Engle used the Eriksen flanker paradigm and not the visual arrays, which is a much easier task. In addition, stimuli in the flanker are generally very close to one another, whereas in the present design distractors in the concentric versions of the task appeared at least 9.3° away from the center of the screen. Perhaps it takes much less time to adjust allocation of attention across the visual field when there is larger visual separation between targets and distractors, consistent with the zoom-lens model offered in Eriksen and Yeh (1985) and used by Heitz and Engle to interpret their findings.

As for why performance on the longest (700 ms) cue-to-stimulus trials was overall most predictive, this is not immediately clear. One possibility that is counter to the previous explanations is that because the present paradigms were much more difficult than the typical flanker tasks, even higher-ability individuals needed longer than the 100 or 400 ms afforded to them to prepare for the trial, and 700 ms is beginning to hit a sweet spot for preparatory time for individual differences to emerge. This might also be because Heitz and Engle (2007) were concerned specifically with spotlight attention, whereas the present tasks involved both spotlight trials and donut-shaped configurations, which are qualitatively different. Finally, it again should be noted that factor scores for the different cue-to-stimulus intervals were strongly correlated (short and long intervals were correlated *r* = .85), and so the results from the semipartial analyses are likely to be relatively unstable, similar to the set sizes comparisons from Table 5.

Table 11 – K Scores as a Function of Cue-to-Stimulus Interval.

	100 ms	400 ms	700 ms	Cohen's d (100 vs. 700)
CVA	2.31	2.36	2.46	0.23*
ECVA	1.84	1.96	1.91	0.10 (*ns*)

Table 12 – Semipartial Correlations of Different Cue-to-Stimulus Intervals.

Construct	Short	Medium	Long	Total R
Attention Control	.13*	.04	.18*	.66*
Fluid Intelligence	.00	.22*	.13*	.63*
Working Memory Capacity	-.02	.02	.28*	.57*
Multitasking	.09	.12	.12	.59*

Note. Short = 100 ms; cue-to-stimulus interval; medium = 400 ms; long = 700 ms. All variables shown are factor scores except multitasking, which is a z-score composite of two tasks. Short and medium factor scores correlated $r = .79$, short and long $r = .85$, and medium and long $r = .81$. $*p < .05$.

CHAPTER 3. GENERAL DISCUSSION

This study was designed in part to be a replication and follow-up of the toolbox approach for assessing individual differences (Draheim et al., 2020b). In this regard, it was both a success and a failure. On the successful side, I found that multiple visual arrays task variants were psychometrically strong indicators of attention control, stronger even than VA4 which was one of the strongest attention control measures in Draheim et al. But, this study failed to replicate the novel finding from Draheim et al. that attention control fully accounted for the WMC-Gf relationship. This challenges our lab's argument that individual differences in the ability to control attention should explain most, if not all, individual differences in executive functioning more generally (e.g., Burgoyne & Engle, 2020; Draheim et al., 2021a; 2021b; Mashburn et al., 2020). Further, although the SACT and Stroop DL tasks performed very well in the present dataset, the flanker DL task appeared worse than in Draheim et al., despite the changes we made to it which were supposed to be improvements. These matters will need to be explored further when data collection for the study at large is completed, which will add around 200 additional participants to the dataset.

One of the overarching findings from this study was that the visual arrays paradigm is remarkably robust to a variety of manipulations. *It is really hard to mess this task up*. I varied whether or not selection was required, what type of selection, the spatial configuration of the stimuli, the number of targets, the number of distractors, and cue-to-stimulus duration, and yet performance in the various tasks and trial types correlated very strongly at both the task- and factor score-level, loaded onto the same factor in latent analyses, and had *roughly* the same predictive validity to other cognitive measures. This finding is of scientific interest because many paradigms require careful administration, meticulous piloting, careful selection of parameters, etc., in order to find the effects of interest. Not so with the visual arrays task. Of note is that most manipulations did produce a moderate-to-large difference in performance (i.e., they made the tasks easier or more difficult), but the manipulations did not change the *nature* of the task. This finding is highly consistent with Shipstead et al. (2014) who wrote that, "selective filtering requirements introduce certain attention control demands that are not reflected in standard visual arrays performance. Nonetheless, all visual arrays tasks have a particularly strong relationship to attention control, regardless of specific demands" (p. 120). To that end, I found that the non-selective visual arrays task here was much less correlated to other cognitive measures and loaded much more weakly with the other attention control and visual arrays tasks, but that it still correlated more strongly with attention control than WMC or Gf, indicating individual differences in performance in even non-selective visual arrays is primarily attentional in nature.

At the task level, the concentric selective visual arrays task with a 1:1 distractor to target ratio and the color selective visual task with a 1:1 distractor ratio correlated very strongly (r = .80). But the concentric task had a clear advantage in terms of its relationship with other attention control measures, Gf, and WMC. This is an encouraging finding in that it suggests location-based selection in the visual arrays paradigm can be as viable, if not more so, for individual differences research than a more typical feature-based (color) selection version. It is also interesting to note that within the CVA task the trial types which were the most predictive were set size 3 of both the spotlight and donut-shaped array configurations. The finding that individual differences in set size 3 performance on relatively easy spotlight trials are as strongly predictive of cognition as any other trial type from any visual arrays task included here is noteworthy, as it suggests that the visual arrays paradigm does not need to be overly complex or difficult in order for important individual differences to emerge. In fact, spotlight trial performance on CVA set size 3 was much more predictive overall than performance from any set size and trial type in the enhanced version of this task, indicating that simpler may be better when it comes to the visual arrays paradigm. This was also supported by the finding that the most complex and difficult task, ECVA, was the least predictive among the selective variants. Regarding the enhanced color selective visual arrays (EVA4) task, it correlated to other cognitive abilities a bit more so than the standard color selective visual arrays (VA4) task, which correlated to other cognitive abilities more so than the non-selective task (NSVA), indicating that both the presence and number of distractors produced more meaningful individual differences in task performance.

Interestingly, most manipulations to the visual arrays tasks resulted in moderate-to-large changes in difficulty, as evidenced by a decrease in *k* scores and accuracy rates. However, this did not always result in a substantive change in the relationship to attention control or other cognitive abilities, and factor score correlations were typically very strong among the various trial types. Zero-order correlations comparing different trial types usually did not produce statistically significant differences in those correlations to other cognitive variables. Using regression analyses, in many instances semipartial correlations to various criterion measures were stronger for one manipulation over another, but the pattern of results was not always consistent or easy to interpret – potentially because of the relative instability of the semipartial correlations given the strong relationship among the factor scores for the different trial types and the relative lack of power in the study due to a smaller listwise set size than hopes. It will be interesting to see if these results become more clear when all the data are collected.

Finally, this study provided insights regarding visual arrays as a measure of attention control. Selective visual arrays in particular loaded strongly with other attention control tasks at the latent level, but it was possible to separate out factors for visual arrays performance and performance on other attention control measures. These factors were very strongly correlated but shared unique variance to other cognitive abilities and criterion measures above and beyond the other in regression analyses, although only the visual arrays factor predicted unique variance to Gf and WMC in structural equation models. These results show that the selective visual arrays paradigm is a particular strong measure of attention control and that there is meaningful variance in the selective visual arrays paradigm not present in the other attention control measures included here.

CHAPTER 4. CONCLUSION

In the present study, I manipulated a variety of aspects of the visual arrays paradigm in an effort to better understand the nature of individual differences in attention control. Results showed most manipulations altered the difficulty of the task but not the nature of what the task measures. Consistent with Draheim et al. (2021b) and Martin et al. (in press), selective visual arrays is a particularly strong and robust indicator of attention control. The aspects of this paradigm that appeared most sensitive to individual differences in attention control were whether distractors were present (i.e., is the task selective) and the ratio of distractors to targets. Beyond that, configuring the stimuli in a concentric manner and requiring location-based selection was found to increase meaningful individual variation in performance over and above color-based selection for simpler versions of the paradigm, but the opposite was true for versions with more distractors. Stronger conclusions can be made when data collection for the study-at-large is completed.

APPENDIX A: ADDITIONAL INFORMATION

Table A1 – Full List of Session 1-4 Tasks in the Study at Large.

Session 1	(Time)	Session 2	(Time)	Session 3	(Time)	Session 4	(Time)
Demographics	3	Color Selective VA	15	Enh. Color Selective VA	15	MicroDyn Complex Problem Solving	45
Non-Selective VA	15	Concentric VA	15	Enh. Concentric VA	15	Running Digit	10
SymSpan	25	Running Letter	10	RotSpan	25	Paper Folding	7.5
RAPM	10	LetterSets	7	NumberSeries	5	Mental Counters	10
SynWin	20	Foster MT Task	20	Auditory FlankerDL	10	ControlTower	20
Double Flanker	3	Visual FlankerDL	15	Auditory StroopDL	10	Hick Task	10
Double Stroop	3	Visual StroopDL	15	Auditory Simon	10	Double Flanker	3
Double Simon	3	IT_2L2D	13	Antisaccade	10	Double Stroop	3
SACT	20	Adaptive Prosaccade	10	IT_4L	12	Double Simon	3
Prosaccade	10	Visual Search Hands	7	Multiple Object Tracking	12.5	Pattern Comparison	3
IT_Standard	13					Letter/Number Comparison	3
Visual Search Lett	7					Digit/Symbol Comparison	3
						Visual Search Arrows	7

Note. Time column are estimates and not calculated from real data from the study. IT = inspection time; DL = deadline.

Table A2 – Zero-Order Correlations Among Visual Arrays Tasks and Other Constructs.

Measure	AC	Gf	WMC	MTz	PS	Avg.
Nonselective	.49	.44	.44	.35	.45	*.43*
Color Selective	.57	.49	.48	.48	.40	*.48*
Enhanced Color Selective	.59	.56	.52	.54	.32	*.51*
Concentric	.63	.54	.53	.49	.45	*.53*
Enhanced Concentric	.53	.45	.39	.46	.33	*.43*
Average	*.56*	*.50*	*.47*	*.46*	*.39*	*.48*

Note. AC = attention control factor scores (antisaccade, SACT, Stroop DL); Gf = fluid intelligence factor scores (Raven, letter sets, number series); WMC = working memory capacity factor scores (rotation span and symmetry span), MTz = multitasking z-score composite (Foster task and synthetic work), PS = processing speed factor scores (digit, letter, and pattern string comparisons); Avg. = average correlation for that VA task; Average = average correlation for that outcome. All correlations are statistically significant at the .05 level.

REFERENCES

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: The same or different constructs? *Psychological Bulletin, 131,* 30-60.

Anllo-Vento, L., & Hillyard, S. A. (1996). Selective attention to the color and direction of moving stimuli: Electrophysiological correlates of hierarchical feature selection. *Perception & Psychophysics, 58*(2), 191-206.

Baddeley, A. D., & Hitch, G. (1974). Working memory. The Psychology of Learning and Motivation, 8, 47–89.

Beck, D. M., & Lavie, N. (2005). Look here but ignore what you see: effects of distractors at fixation. *Journal of Experimental Psychology: Human Perception and Performance, 31*(3), 592-607.

Bleckley, M. K., Durso, F. T., Crutchfield, J. M., Engle, R. W., & Khanna, M. M. (2003). Individual differences in working memory capacity predict visual attention allocation. *Psychonomic Bulletin & Review, 10*(4), 884-889.

Bleckley, M. K., Foster, J. L., & Engle, R. W. (2015). Working memory capacity accounts for the ability to switch between object-based and location-based allocation of visual attention. *Memory & Cognition, 43*(3), 379-388.

Botvinick, M. M., Cohen, J. D., & Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: An update. *Trends in Cognitive Sciences, 8*(12), 539-546.

Broadway, J.M., Redick, T.S., & Engle, R.W. (2010). Individual differences in working memory capacity: Control is (in) the goal. In R. Hassin, K. Ochsner, & Y.Trope (Eds.). *Self-control in society, mind, and brain.* Oxford: Oxford University Press. (pp 163 – 174).

Cave, K. R., & Bichot, N. P. (1999). Visuospatial attention: Beyond a spotlight model. *Psychonomic Bulletin & Review, 6*(2), 204-223.

Colom, R., & Shih, P. C. (2004). Is working memory fractionated onto different components of intelligence? A reply to Mackintosh and Bennett (2003). *Intelligence, 32*(5), 431-444. https://doi.org/10.1016/j.intell.2004.06.011

Comrey, A. L., & Lee, H. B. (1992). Interpretation and application of factor analytic results. In A. L. Comrey, & H. B. Lee (Eds). *A first course in factor analysis.*

Conway, A. R. A., Cowan, N., Bunting, M. F., Therriault, D. J., & Minkoff, S. R. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence, 30*, 163183.

Conway, A. R. A., Kovacs, K., Hao, H., Rosales, K. P., & Snijder, J-P. (2021). Individual differences in attention and intelligence: A united cognitive/psychometric approach. *Journal of Intelligence, 9*(3), 34. https://doi.org/10.3390/jintelligence9030034

Cowan, N., Elliott, E. M., Saults, J. S., Morey, C. C., Mattox, S., Hismjatullina, A., & Conway, A. R. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive psychology*, *51*(1), 42-100.

Cowan, N., & Morey, C. C. (2006). Visual working memory depends on attentional filtering. *Trends in Cognitive Sciences, 10*(4), 139-141.

Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist, 12*(11), 671-684. https://doi.org/10.1037/h0043943

Cronbach, L. J., & Furby, L. (1970). How should we measure "change"—or should we?. *Psychological Bulletin*, *74*(1), 68-80.

Davies, D. R., Jones, D. M., & Taylor, A. (1984). Selective- and sustained-attention tasks: Individual and group differences. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of attention.* (pp. 396-447). Academic Press.

DiStefano, C., Zhu, M., & Mindrilã, D. (2009). Understanding and using factor scores: Considerations for the applied researcher. *Practical Assessment, Research, and Evaluation*, *14*(1), Article 20.

Draheim, C., Harrison, T. L., Embretson, S. E., & Engle, R. W. (2018). What item response theory can tell us about the complex span tasks. *Psychological Assessment*, *30*(1), 116-129.

Draheim, C., Hicks, K. L., & Engle, R. W. (2016). Combining reaction time and accuracy: The relationship between working memory capacity and task-switching as a case example. *Perspectives on Psychological Science*, *11*(1), 133-155.

Draheim, C., Mashburn, C. A., Martin, J. D., & Engle, R. W. (2019). Reaction time in differential and development research: A review and commentary on problems and alternatives. *Psychological Bulletin*, *145*(5), 508-535.

Draheim, C., Pak, R., Draheim, A. A., & Engle, R. W. (2021, August 5). The role of attention control in complex real-world tasks. *PsyArXiv*, https://doi.org/10.31234/osf.io/9ekpu

Draheim, C., Tsukahara, J. S., Martin, J. D., Mashburn, C. A., & Engle, R. W. (2021b). A toolbox approach to improving the measurement of attention control. *Journal of Experimental Psychology: General, 150*(2), 242–275. https://doi.org/10.1037/xge0000783

Driver, J., & Baylis, G. C. (1989). Movement and visual attention: the spotlight metaphor breaks down. *Journal of Experimental Psychology: Human Perception and Performance, 15*(3), 448-456.

Egeth, H. E., & Yantis, S. (1997). Visual attention: control, representation, and time course. *Annual Review of Psychology, 48*, 269–297.

Egly, R., Driver, J., & Rafal, R. D. (1994). Shifting visual attention between objects and locations: Evidence from normal and parietal lesion participants. *Journal of Experimental Psychology: General, 123*, 161–177.

Egly, R., & Homa, D. (1984). Sensitization of the visual field. *Journal of Experimental Psychology: Human Perception and Performance*, *10*, 778–793.

Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). *Manual for kit of factor referenced cognitive tests*.: Educational testing service.

Elsmore, T. F. (1994). SYNWORK: APC-based tool for assessment of performance in a simulated work environment. *Behavior Research Methods, Instrumentation, and Computers*, 26, 421-426. http://dx.doi.org/10.3758/BF03204659

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

Engle, R. W. (2018). Working memory and executive attention: A revisit. *Perspectives on Psychological Science, 13*(2), 190-193.

Engle, R. W., & Kane, M. J. (2004). Executive attention, working memory capacity, and a two-factor theory of cognitive control. In B. Ross (Ed.), *The Psychology of Learning and Motivation* (Vol. 44, pp. 145-199). NY: Elsevier.

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

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics, 16*(1), 143-149.

Eriksen, C. W., & James, J. D. S. (1986). Visual attention within and around the field of focal attention: A zoom lens model. *Perception & Psychophysics, 40*(4), 225-240.

Eriksen, C. W., & Yeh, Y.-Y. (1985). Allocation of attention in the visual field. *Journal of Experimental Psychology: Human Perception and Performance*, *11*(5), 583–597

Fan, J., McCandliss, B. D., Sommer, T., Raz, A., & Posner, M. I. (2002). Testing the efficiency and independence of attentional networks. *Journal of Cognitive Neuroscience*, *14*(3), 340-347.

Feldt, L. S. (1961). The use of extreme groups to test for the presence of a relationship. *Psychometrika, 26*, 307–316.

Forstmann, B. U., Tittgemeyer, M., Wagenmakers, E. J., Derrfuss, J., Imperati, D., & Brown, S. (2011). The speed-accuracy tradeoff in the elderly brain: a structural model-based approach. *Journal of Neuroscience, 31*(47), 17242-17249.

Friedman, N. P., & Miyake, A. (2004). The relations among inhibition and interference control functions: a latent-variable analysis. *Journal of Experimental Psychology: General*, *133(1),* 101-135.

Friedman, N. P., & Miyake, A. (2017). Unity and diversity of executive functions: Individual differences as a window on cognitive structure. *Cortex, 86*, 186-204.

Fukuda, K., & Vogel, E. K. (2009). Human variation in overriding attentional capture. *Journal of Neuroscience, 29*(27), 8726-8733.

Fukuda, K., & Vogel, E. K. (2011). Individual differences in recovery time from attentional capture. *Psychological science*, *22*(3), 361-368.

Fukuda, K., Woodman, G. F., & Vogel, E. K. (2015). Individual differences in visual working memory capacity: Contributions of attentional control to storage. *Mechanisms of Sensory Working Memory: Attention and Performance XXV*, 105.

Goodhew, S. C., & Edwards, M. (2019). Translating experimental paradigms into individual differences research: Contributions, challenges, and practical recommendations. *Consciousness and Cognition*, *69*, 14-25.

Hedge, C., Powell, G., Bompas, A., & Sumner, P. (in press). Strategy and processing speed eclipse individual differences in control ability in conflict tasks. *Journal of Experimental Psychology: Learning, Memory and Cognition*. https://doi.org/10.31234/osf.io/vgpxq

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods, 50*(3), 1166-1186.

Hedge, C., Vivian-Griffiths, S., Powell, G., Bompas, A., & Sumner, P. (2019). Slow and steady? Strategic adjustments in response caution are moderately reliable and correlate across tasks. *Consciousness and Cognition, 75*, 102797. https://doi.org/10.1016/j.concog.2019.102797

Heitz, R. P. (2014). The speed-accuracy tradeoff: History, physiology, methodology, and behavior. *Frontiers in Neuroscience, 8*.

Hertzog, C., Vernon, M. C., & Rypma, B. (1993). Age differences in mental rotation task performance: The influence of speed/accuracy tradeoffs. *Journal of Gerontology, 48*(3), 150-156.

Huettel, S. A., Güzeldere, G., & McCarthy, G. (2001). Dissociating the neural mechanisms of visual attention in change detection using functional MRI. *Journal of Cognitive Neuroscience, 13*(7), 1006-1018.

Hughes, M. M., Linck, J. A., Bowles, A. R., Koeth, J. T., & Bunting, M. F. (2014). Alternatives to switch-cost scoring in the task switching paradigm: Their reliability and increased validity. *Behavioral Research Methods, 46*(3), 702-721.

Hutchison, K. A. (2007). Attentional control and the relatedness proportion effect in semantic priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(4), 645-662.

Jonides, J. (1983). Further toward a model of the mind's eye's movement. *Bulletin of the Psychonomic Society*, *21*(4), 247-250.

Kaernbach, C. (1991). Simple adaptive testing with the weighted up-down method. *Perception & Psychophysics, 49*(3), 227–229.

Kane, M. J., Bleckley, M. K., Conway, A. R., & Engle, R. W. (2001). A controlledattention view of working-memory capacity. *Journal of Experimental Psychology: General, 130*, 169-183.

Kane, M. J., Conway, A. R., Hambrick, D. Z., & Engle, R. W. (2007). Variation in working memory capacity as variation in executive attention and control. *Variation in working memory*, *1*, 21-48.

Kane, M. J., & Engle, R. W. (2003). Working memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General*, *132*, 47-70.

Kane, M. J., Hambrick, D. Z., & Conway, A. R. A. (2005). Working memory capacity and fluid intelligence are strongly related constructs: comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin, 131*(1), 66–71.

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: a latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General*, *133*(2), 189-217.

Kane, M. J., & McVay, J. C. (2012). What mind wandering reveals about executive-control abilities and failures. *Current Directions in Psychological Science*, *21*(5), 348-354.

Kane, M. J., Meier, M. E., Smeekens, B. A., Gross, G. M., Chun, C. A., Silvia, P. J., & Kwapil, T. R. (2016). Individual differences in the executive control of attention, memory, and thought, and their associations with schizotypy. *Journal of Experimental Psychology: General*, *145*(8), 1017-1048.

Kovacs, K., and Conway, A. R. A. (2016). Process overlap theory: A unified account of the general factor of intelligence. *Psychological Inquiry, 27*(3), 151-177. https://doi.org/10.1080/1047840X.2016.1153946

LaBerge, D. (1983). Spatial extent of attention to letters and words. *Journal of Experimental Psychology: Human Perception and Performance, 9*(3), 371-379.

Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology: Human Perception and Performance, 21*(3), 451-468.

Logie, R. H., Della Sala, S., Laiacona, M., Chalmers, P., & Wynn, V. (1996). Group aggregates and individual reliability: The case of verbal short-term memory. *Memory & Cognition*, *24*(3), 305-321.

Lord, F. M. (1956). The measurement of growth. *ETS Research Bulletin Series*, *1956*(1), i-22.

Luce, R.D. (1986). *Response times: Their role in inferring elementary mental organization.* Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195070019.001.0001

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390*, 279–281.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin, 109*(2), 163-203.

Martin, J. D., Mashburn, C. A., & Engle, R. W. (2020). Improving the validity of the Armed Service Vocational Aptitude Battery with measures of attention control. *Journal of Applied Research in Memory and Cognition*, *9*(3), 323-335. https://doi.org/10.1016/j.jarmac.2020.04.002

Martin, J. D., Tsukahara, J. S., Draheim, C., Shipstead, Z. Mashburn, C. A., & Engle, R.

W. (in press). The visual arrays task: Visual working memory capacity or attention control? *Journal of Experimental Psychology*. https://psyarxiv.com/u92cm/

Mashburn, C. M., Tsukahara, J. T., & Engle, R. W. (2020). Individual differences in attention control: Implications for the relationship between working memory capacity and fluid intelligence. In R. Logie, V. Camos, N. Cowan (Eds.), Working Memory: *The state of the science* (pp. 175-211). Oxford University Press. https://doi.org/10.1093/oso/9780198842286.003.0007

Matsunaga, M. (2010). How to factor-analyze your data right: do's, don'ts, and how-to's. *International Journal of Psychological Research, 3*(1), 97-110.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience, 24*(1), 167-202.

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science, 21*(1), 8-14.

Müller, M. M., & Hübner, R. (2002). Can the spotlight of attention be shaped like a doughnut? Evidence from steady-state visual evoked potentials. *Psychological Science, 13*(2), 119-124.

Müller, M. M., Malinowski, P., Gruber, T., & Hillyard, S. A. (2003). Sustained division of the attentional spotlight. *Nature*, *424*(6946), 309-312.

Norman, D. A., & Shallice, T. (1986). Attention to action. Willed and automatic control of behavior. In Consciousness and Self-Regulation (pp. 1–18). Springer.

Nunnally, J. C. (1964). *Educational measurement and evaluation*. McGraw-Hill.

Oberauer, K., Schulze, R., Wilhelm, O., & Süß, H.M. (2005). Working memory and intelligence—their correlation and their relation: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin, 131*, 66-71.

Paap, K. R., & Sawi, O. (2016). The role of test-retest reliability in measuring individual and group differences in executive functioning. *Journal of Neuroscience Methods, 274*, 81-93.

Parsons, S., Kruijt, A. W., & Fox, E. (2019). Psychological science needs a standard practice of reporting the reliability of cognitive-behavioral measurements. *Advances in Methods and Practices in Psychological Science, 2*(4), 378-395.

Posner, M. I., & DiGirolamo, G. J. (1998). Conflict, target detection and cognitive control. in R. Parasuraman (Ed.), The attentive brain (pp. 401-423). MIT Press.

Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience, 13*(1), 25-42.

Posner, M. I., Snyder, C. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General, 109*(2), 160-174.

Preacher, K. J. (2015). Extreme groups designs. In R. L. Cautin & S. O. Lilienfeld (Eds.), *The encyclopedia of clinical psychology* (Vol. 2, pp. 1189-1192). John Wiley & Sons Inc.

Raven, J. C., & J. H. Court. (1998). *Raven's progressive matrices and vocabulary scales.* Oxford, UK: Oxford Psychologists Press.

Redick, T. S., Shipstead, Z., Meier, M. E., Montroy, J. J., Hicks, K. L., Unsworth, N., ... & Engle, R. W. (2016). Cognitive predictors of a common multitasking ability: Contributions from working memory, attention control, and fluid intelligence. *Journal of Experimental Psychology: General, 145*(11), 1473.

Rensink, R. A., O'regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science, 8*(5), 368-373.

Rey-Mermet, A., Gade, M., & Oberauer, K. (2018). Should we stop thinking about inhibition? Searching for individual and age differences in inhibition ability. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44*(4), 501-526.

Rey-Mermet, A., Gade, M., Souza, A. S., von Bastian, C. C., & Oberauer, K. (2019). Is executive control related to working memory capacity and fluid intelligence?. *Journal of Experimental Psychology: General, 148*(8), 1335-1372.

Rouder, J. N., & Haaf, J. M. (2019). A psychometrics of individual differences in experimental tasks. *Psychonomic Bulletin & Review*, 26(3), 772-789.

Rouder, J., Kumar, A., & Haaf, J. M. (2019, March 25). Why Most Studies of Individual Differences With Inhibition Tasks Are Bound To Fail. PsyArXiv: https://doi.org/10.31234/osf.io/3cjr5

Rueda, M. R. (2018). Attention in the heart of intelligence. *Trends in Neuroscience andEducation, 13*, 26-33. https://doi.org/10.1016/j.tine.2018.11.003

Schmeichel, B. J., & Demaree, H. A. (2010). Working memory capacity and spontaneous emotion regulation: High capacity predicts self-enhancement in response to negative feedback. *Emotion*, *10*(5), 739-734.

Shipstead, Z., & Engle, R. W. (2013). Interference within the focus of attention: Working memory tasks reflect more than temporary maintenance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(1), 277-289.

Shipstead, Z., Harrison, T. L., & Engle. (2015). Working memory capacity and the scope and control of attention. *Attention, Perception, & Psychophysics*, *77*. 1863-1880.

Shipstead, Z., Harrison, T. L., & Engle. (2016). Working memory capacity and fluid intelligence: Maintenance and disengagement. *Perspectives on Psychological Science, 11*(6), 771-799.

Shipstead, Z., Lindsey, D. R., Marshall, R. L., & Engle, R. W. (2014). The mechanisms of working memory capacity: Primary memory, secondary memory, and attention control. *Journal of Memory and Language, 72*(1), 116-141. https://doi.org/10.1016/j.jml.2014.01.004

Shipstead, Z., Redick, T. S., Hicks, K. L., & Engle, R. W. (2012). The scope and control of attention as separate aspects of working memory. *Memory, 20*(6), 608-628.

Starns, J. J., & Ratcliff, R. (2010). The effects of aging on the speed-accuracy compromise: Boundary optimality and the diffusion model. *Psychology and Aging, 25*(2), 377390.

Stevens, J.P. (1992). Applied Multivariate Statistics for the Social Sciences (2nd edition). Hillsdale, NJ: Erlbaum.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*, 643-662.

Thurstone, L. L. (1938). *Primary Mental Abilities.* Chicago, IL: University of Chicago Press.

Tsukahara, J. S., Harrison, T. L., Draheim, C., Martin, J. D., & Engle, R. W. (2020). Attention control: The missing link between sensory discrimination and intelligence. *Attention, Perception, and Psychophysics, 82*(7), 3445-3478. https://doi.org/10.3758/s13414-020-02044-9

Ulrich, R., Schröter, H., Leuthold, H., & Birngruber, T. (2015). Automatic and controlled stimulus processing in conflict tasks: Superimposed diffusion processes and delta functions. *Cognitive Psychology, 78,* 148-174.

Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: active maintenance in primary memory and controlled search from secondary memory. *Psychological Review, 114*(1), 104-132.

Unsworth, N., Fukuda, K., Awh, E., & Vogel, E. K. (2014). Working memory and fluid intelligence: Capacity, attention control, and secondary memory retrieval. *Cognitive Psychology, 71,* 1-26.

Unsworth, N., & Miller, A. L. (2021). Individual Differences in the Intensity and Consistency of Attention. *Current Directions in Psychological Science.* https://doi.org/10.1177/09637214211030266

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

Unsworth, N., Redick, T. S., Heitz, R. P., Broadway, J. M., & Engle, R. W. (2009). Complex working memory span tasks and higher-order cognition: A latent-variable analysis of the relationship between processing and storage. *Memory, 17(6)*, 635654.

Unsworth, N., & Spillers, G. J. (2010). Working memory capacity: Attention control, secondary memory, or both? A direct test of the dual-component model. *Journal of Memory and Language*, *62*(4), 392-406.

Valdes-Sosa, M., Bobes, M. A., Rodriguez, V., & Pinilla, T. (1998). Switching attention without shifting the spotlight: Object-based attentional modulation of brain potentials. *Journal of Cognitive Neuroscience, 10*(1), 137-151.

Verhaeghen, P., & De Meersman, L. (1998). Aging and the Stroop effect: A meta-analysis. *Psychology and Aging*, *13*(1), 120-126.

Vogel, E. K., McCollough, A. W., & Machizawa, M. G. (2005). Neural measures reveal individual differences in controlling access to working memory. *Nature*, *438*(7067), 500-503.

Von Bastian, C. C., Blais, C., Brewer, G. A., Gyurkovics, M., Hedge, C., Kalamala, P., Meier, M. E., Oberauer, K., Rey-Mermet, A., Rouder, J. N., Souza, A. S., Bartsch, A. S., Conway, A. R. A., Draheim. C., Engle, R. W., Fischkorn, G. T., Friedman, N. P., Gustavson, D. E., Koch, I., . . . Whitehead, P. S. (2020, July 27). Advancing the understanding of individual differences in attentional control: Theoretical, methodological, and analytical considerations. *PsyArXiv*. https://doi.org/10.31234/osf.io/x3b9k

Whitehead, P. S., Brewer, G. A., & Blais, C. (2020). Reliability and Convergence of Conflict Effects. *Experimental Psychology*, *67*(5), 303-313.

Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta Psychologica*, *41*(1), 67-85.