UNIVERSITY OF PADOVA

Department of General Psychology

Master Degree in Cognitive Neuroscience and Clinical Neuropsychology

Final dissertation

Lost in Estimation: Untangling the Enigma between Numerical Acuity and Visual Working Memory

Supervisor
Dr. Roberto Dell'Acqua

Co-supervisor
Dr. Mattia Doro

Candidate: Maaria Vilhelmiina Talvio

Student ID number: 2081318

Academic Year 2023/2024

Abstract

In daily life, humans – much like a wide variety of animal species – constantly make comparisons, such as choosing the shortest line at the store or selecting the fullest basket of strawberries. Yet, the cognitive resources that enable the ability to discriminate large numerical quantities using only visual information, and without relying on explicit counting, remain unclear. Prior studies suggest that performance on numerosity comparison tasks partly relies on domain-general visual working memory (VWM) resources, facilitating us to selectively attend to numerical information while suppressing non-numerical, potentially distracting, visual information. In this framework, the present thesis explores whether known limits in VWM capacity place any constraints on the precision of numerosity discrimination, also known as numerical acuity. To this end, forty-six university students completed a numerosity comparison task measuring numerical acuity alongside a prototypical change detection task measuring VWM capacity. The results showed that in spite of rigorous visual controls, non-numerical visual cues significantly affected the accuracy of numerosity discrimination. However, no direct inter-individual correlation was found between VWM capacity and numerical acuity, nor between VWM capacity and variation in numerical acuity across trials characterized by distinct non-numerical visual covariates. These findings suggest that while non-numerical visual covariates undoubtedly influence our ability to discriminate large numerosities, numerical acuity likely relies on a processing stage that is functionally independent from VWM. Future studies utilizing direct neural measures of VWM capacity, including its attention-based filtering efficiency, are warranted to substantiate these findings.

Keywords: approximate number system, visual working memory, numerical acuity, correlation, arithmetic

Contents

1 Introduction

We continuously engage in comparisons within our visual surroundings. For instance, we question whether we picked the shortest line to the cashier at the grocery store, ponder over which of two fruit cups on the cafeteria shelf contain more blueberries, or before stepping onto a metro carriage, we assess the number of people in different parts of the metro. Extracting estimates of items in a visual scene is crucial for informing the numerous choices we make in our daily lives. Converging evidence in research indicates that our ability to perceive numerical magnitude is spontaneous and based on a non-verbal system that allows for approximate estimates of large discrete non-symbolic numerical quantities (i.e., numerosities) at a glance without the necessity of counting (Dehaene, 2009; Feigenson et al., 2004). While this ability is widely considered a foundational skill of mathematical learning, the exact perceptual and cognitive mechanisms underlying it continue to be a topic of substantial debate (Dowker, 2023).

For smaller quantities, falling within the range of one to four items, judgments of numerosity tend to be considerably more precise than those for larger numbers (Melcher et al., 2021; Piazza, 2010; Revkin et al., 2008). This rapid and accurate recognition of small quantities without counting, termed subitizing, seems to depend on the presence of domain-general cognitive abilities necessary for information processing across domains, such as working memory and attention (Chen et al., 2022; Pagano et al., 2014; Piazza et al., 2011). A number of studies have demonstrated that both subitizing and visual working memory (henceforth, VWM) share a similar upper-limit capacity restricted to about four items, and performance in these two tasks often correlate (Ashkenazi et al., 2022; Formoso et al., 2017; Luck & Vogel, 1997; Piazza et al., 2011). This mutual influence has been interpreted to suggest that both subitizing and VWM relies on a shared processing stage underlying object individuation, which is the ability of the perceptual system to select a fixed number (around four) of visual objects from a crowded scene based on their spatial information (Drew & Vogel, 2008; Hyde, 2011; Piazza et al., 2011; Xu & Chun, 2009). This so-called parallel individuation of multiple objects at a time has been hypothesized to correspond to an attentional priority map or salience map of the salient object locations (Knops et al., 2014; Melcher & Piazza, 2011; Piazza et al., 2011). In such a map, some items receive preferential processing by bottom-up (e.g., size) or top-down (e.g., task relevance) factors, with flexible capacity limits determined by competitive interactions between the items (Fecteau & Munoz, 2006; Knops et al., 2014). In accordance with the salience map theories, enhancing the relative salience of one item in a set has been shown to decrease both the subitizing range and VWM capacity for all the other less salient items (Melcher & Piazza, 2011). Thus, suggesting that the relative salience, rather than just the allocation of attention itself, contributes to determining the range of subitizing.

With larger quantities, the extent to which domain-general cognitive resources related to working memory are involved remains debated. Research has established that when dealing with numerosities beyond the subitizing range, judgments rely on estimation, which, unlike subitizing, tends to be imprecise, with a variability that typically increases as the number of objects is increased (Jevons, 1871). This dichotomy between subitizing and estimation has given rise to the idea that these processes may be supported, at least in part, by distinct mechanisms (Feigenson et al., 2004; Revkin et al., 2008). Numerosity estimation has been widely believed to rely on a separate preverbal cognitive system, commonly referred to as the approximate number system (henceforth, ANS), that enables both the estimation and approximate operation of discrete quantity representations (Dehaene, 2009; Feigenson et al., 2004). However, increasing evidence has recently started to suggest that the commonly used paradigm assessing ANS ability, numerosity comparison task, where participants are typically asked to identify the numerically larger of two dot arrays, might not be solely domain-specific, i.e., a system that functions to represent only a particular kind of entity. Instead, the task may also depend on domain-general functions that enable individuals to focus on the dimensions of quantity while disregarding non-numerical aspects of the stimuli (Formoso et al., 2017; Lee & Cho, 2019; Rodríguez & Ferreira, 2023). Indeed, in support of this, perceptual properties unrelated to numerosity, such as dot size and surface area, have been discovered to increase the difficulty of numerosity discrimination for both young children and adults (Gebuis & Reynvoet, 2012a; Szűcs et al., 2013). Thus, suggesting that performance on numerosity comparison tasks may not reflect the ANS but rather a combination of abilities, including domain-general cognitive abilities such as VWM (Lee & Cho, 2019).

The primary objective of the present thesis is to help clarify the functional relationship between the degree of efficiency of discriminating numerosities of varying ratios in a numerosity comparison task, i.e., numerical acuity (Dehaene, 2011), and VWM. Brain systems processing and estimating magnitudes of various sets of quantities are thought to play a crucial role in establishing the foundational elements for a higher-level comprehension of mathematics (Feigenson et al., 2004; Menon & Chang, 2021). Thus, understanding the mechanisms underlying the discrimination of large discrete non-symbolic numerical quantities can provide insight into the fundamental cognitive processes that underpin mathematical reasoning and problem-solving. The initial phase of this thesis will entail a thorough review of the existing literature, encompassing the key research findings concerning both the behavioral manifestations and neural correlates of numerical acuity and VWM. Subsequently, the thesis will proceed to articulate its research question and hypotheses.

1.1 Approximate Number System

In situations when counting is prevented, such as with a brief stimulus presentation, the evaluation of larger numerosities exceeding the subitizing range (i.e., more than four items) relies on numerical estimation, a rapid apprehension of an approximate representation of numerical quantity. The ANS is widely assumed to support not only estimation but also comparison and basic numerical operations, such as approximate arithmetic of numerosities without the need for number symbols or language (Dehaene, 2001b, 2009). Several lines of research have suggested that the ANS is a universal, innate, and separate non-verbal cognitive system (Dehaene, 2001b; Halberda & Feigenson, 2008; Odic & Starr, 2018; Pica et al., 2004), as the ability to discriminate approximate numerical quantities is evident across the lifespan, from newborn infants (Izard et al., 2009; Xu & Spelke, 2000) to highly proficient adults in mathematics (Guillaume et al., 2013). Additionally, this ability is observed in a wide range of animal species, including fish (Agrillo et al., 2010; Potrich et al., 2015), birds (Ditz & Nieder, 2016), and non-human primates (Flombaum et al., 2005).

A commonly used method to investigate individuals' ANS ability is to implement a non-symbolic numerosity comparison task. In this task, participants are typically presented with two dot arrays simultaneously, albeit separately, on a single screen for a very brief duration (typically ranging from 200 to 1500 milliseconds between studies) and are instructed to choose the numerically larger array (De Smedt et al., 2013; Dietrich et al., 2015). The rationale behind it is that performance is assumed to follow Weber's law (Halberda & Feigenson, 2008; Libertus & Brannon, 2009), whereby the difference in intensity needed to discriminate two stimuli is proportional to their objective intensities (Fechner, 1860). This implies that while the ability to estimate numerosity is believed to have no specific upper bound, discrimination between any two numerical quantities is dependent on the ratio between the quantities to be compared; as the ratio approaches 1, comparison becomes more difficult, and accuracy tends towards chance levels (Dietrich et al., 2015; Halberda & Odic, 2015).

The ratio-dependent performance in numerosity comparison task, know as numerical acuity, undergoes a significant enhancement during human development. Even before the age of one, the ability to discriminate numerosities progresses from a ratio of 1:2 (e.g., 8 vs. 16) to 2:3 (e.g., 8 vs. 12; Lipton & Spelke, 2003; Xu & Spelke, 2000), and reaching a ratio of 5:6 by preschool years (Halberda & Feigenson, 2008). The threshold continues to mature gradually to young adulthood until it reaches its peak around the age of 30, where an adult can reliably judge a magnitude difference between ratios such as 9:10 to 10:11 (Halberda et al., 2012; Halberda & Feigenson, 2008). Moreover, significant individual differences in numerical acuity have been observed starting early in life (Halberda et al., 2008, 2012; Libertus & Brannon, 2010). Studies have demonstrated that these inter-individual differences correlate with symbolic mathematical performance both prospectively and retrospectively, as well as at different ages and arithmetic skill levels (Chen & Li, 2014; Fazio et al., 2014; Schneider et al., 2017). Similarly, some studies have found that children with mathematical difficulties and developmental dyscalculia, a disorder characterized by moderate to extreme difficulties in fluent numerical computations, exhibit significantly lower numerical acuity (i.e., require larger numerical difference between the two sets of numerosities) than their typically achieving peers (Decarli et al., 2020; Mazzocco et al., 2011; Mussolin et al., 2010; Piazza et al., 2010). These findings have led researchers to suggest that ANS may serve as a cognitive foundation for the later acquired symbolic mathematic knowledge by providing the basis for understanding numerical symbols and symbolic arithmetic operations (Dehaene & Cohen, 1995; Feigenson et al., 2004; Piazza, 2010).

Consequently, deficiencies in the ANS have been suggested to potentially lead to challenges in achieving proficiency in mathematical abilities (Mazzocco et al., 2011; Piazza et al., 2010).

1.1.1 Modeling and Indexing the Approximate Number System

A signature characteristic of the ANS is an imprecise and noisy representation of magnitudes, or in other words, variability in the representation of specific numerosities (Feigenson et al., 2004; Nieder, 2016). For instance, the representation of the numerosity eight is a random variable with a mean of eight and a normally distributed variance. Moreover, as numerosities increase, representations become progressively less precise. This inherent imprecision of the ANS is often modeled as overlapping Gaussian tuning curves on an internal continuum that showcases an increase in overlap as numerosity increases, either due to logarithmic spacing (i.e., the means increase logarithmically with numerosity while the standard deviation is constant for all numerosities) or linearly increasing the spread (i.e., both means and standard deviations increase linearly with numerosity) of the Gaussian distributions (Feigenson et al., 2004; Lindskog et al., 2013). Although the logarithmic and linear models differ in their assumptions about how the ANS represents magnitude, both offer similar behavioral predictions regarding the ability to discriminate between magnitudes (Dehaene, 2001a; Feigenson et al., 2004). However, at the neural level, evidence from studies employing single-cell recordings with monkeys has presented support for the logarithmically compressed representations of numerosities (Nieder & Merten, 2007; Nieder & Miller, 2003).

To assess the spread of the Gaussian tuning curves (i.e., the underlying numerosity representations), several behavioral indices have been employed, with the four most prevalent being accuracy, numerical ratio effect (NRE) for accuracy, NRE for reaction time, and Weber fractions (*w*; Inglis & Gilmore, 2014). The use of accuracy (aggregated across employed ratios) as a metric is grounded on the notion that higher precision in numerosity representations correlates with better task accuracy overall (Dietrich et al., 2015). The NRE, on the other hand, is utilized due to the overlapping nature of numerosity representations, leading participants to exhibit ratio-dependent performance in non-symbolic numerosity comparison tasks. The NRE captures the increase in error rate and reaction time depending on the numerical distance between the two arrays in a task trial; as the numerical ratio becomes larger/closer to 1, both error rate and reaction times increase (Dietrich et al., 2015; Price et al., 2012; Sekuler & Mierkiewicz, 1977). Lastly, the *w*, the predominantly used index, which makes the theoretical assumption that numerosity discrimination accuracy follows Weber's law and it can be computed using classical decision-making models, such as the signal detection model (Inglis & Gilmore, 2014; Odic & Starr, 2018; Piazza et al., 2004). The *w* quantifies the extent to which two sets of objects must vary for an individual to notice a difference between them reliably. This means that the smaller the *w*, the narrower the Gaussian tuning curves, and the better the precision of the internal numerosity representations (Inglis & Gilmore, 2014).

While it is assumed that all these metrics collectively reflect the precision of the ANS and should thus correlate closely, research using numerosity comparison tasks has revealed significant variations in these correlations (Inglis & Gilmore, 2014; Price et al., 2012). These discrepancies have been suggested to arise from differences in reliability, particularly concerning the NRE, which has exhibited limited correlation with the other ANS indices (Inglis & Gilmore, 2014). A comprehensive review by Dietrich and colleagues (2015) even advised against using NRE due to its low reliability; instead, they recommended *w* and accuracy as the preferable measures. Both the *w* and overall accuracy have demonstrated reasonable convergent construct validity and acceptable reliability (Dietrich et al., 2015; Inglis & Gilmore, 2014; Lindskog et al., 2013; Price et al., 2012), thus appearing as reasonable candidates for assessing ANS ability. However, despite the intricate equations used to calculate the Weber fraction, studies have noted that it does not seem to provide additional information about ANS performance beyond the simple accuracy metric, particularly when taking a correlational perspective (Guillaume & Van Rinsveld, 2018; Inglis & Gilmore, 2014). Nevertheless, Guillaume and Van Rinsveld (2018) suggest that an exception for computing *w* may provide a precise psychophysiological modeling study investigating the specific contributions of numerical and non-numerical dimensions to human behavior.

1.2 Neural Substrate of Numerosities

Transitioning from behavioral methods to employing neurophysiological techniques, research has uncovered the presence of numerosity-tuned neurons, which exhibit maximal responses to specific numerosities. These neurons' responsiveness gradually decreases as the numerosity deviates from their specific preference, thus demonstrating imprecision with a bell-shaped Gaussian response function (Nieder, 2016). Initially identified through single-cell recordings in the association cortices of nonhuman primates (Nieder et al., 2002; Nieder & Miller, 2004), neurons specialized in processing visual numerosity have more recently been discovered in humans, particularly in the medial temporal lobe area (Kutter et al., 2018, 2023). However, human studies employing single-cell recordings are scarce. Instead, a more commonly used method, functional magnetic resonance imaging (fMRI), has provided converging evidence of visual numerosity tuning in humans (Tsouli et al., 2022). For instance, fMRI studies utilizing population receptive field modeling approaches (Harvey et al., 2013) and fMRI adaptation

(Piazza et al., 2004) have identified numerosity-tuned neuronal populations in specific areas of the posterior parietal cortex, particularly in the bilateral intraparietal sulcus (IPS), compatible with that observed in macaque monkeys. Moreover, recent fMRI investigations have revealed that various brain regions engaged in processing numerosity, including areas in the parietal, frontal, and occipito-temporal cortex, exhibit a topographical organization with numerosity preference gradually changing along the cortical surface (Cai et al., 2021; Harvey et al., 2013; Harvey & Dumoulin, 2017).

While the precise localization of numerosity processing remains a topic of debate, collective evidence points to a cortical "number network" rather than a direct mapping of number abilities to a single brain region (Harvey & Dumoulin, 2017; Nieder, 2016; Tsouli et al., 2022; Wilkey & Ansari, 2020). Despite, some authors presenting evidence suggesting rapid and direct encoding of numerosity at early stages of the visual processing stream within the occipital cortex, mirroring the processing of other non-numerical visual properties such as shape, color, and size (Castaldi et al., 2019; DeWind et al., 2019; Park et al., 2016; Van Rinsveld et al., 2020), the majority of neuroimaging studies seem to suggest that the key brain regions for numerosities lie in higher levels of information processing, particularly within the extensive network of frontoparietal areas (e.g., Nieder, 2016; Piazza et al., 2007; Sokolowski et al., 2017). Neuropsychological research has also provided support for the higher-level representation of numerosities by indicating that patients with damage to the posterior parietal lobe exhibit difficulties in both symbolic number tasks and non-symbolic numerical estimation (Ashkenazi et al., 2008; Lemer et al., 2003). In addition, consistent with neuropsychological lesion findings, it is theorized that developmental dyscalculia arises from underdeveloped neuronal circuitries within the number network (Butterworth et al., 2011). Notably, studies have linked dyscalculia to reduced gray and white matter volumes within the frontoparietal network. For instance, Rotzer and colleagues (2008) demonstrated that individuals with dyscalculia display reduced gray matter volume in the right IPS, the left inferior frontal gyrus, the anterior cingulum and the bilateral middle frontal gyri, as well as significantly less white matter volume in the left frontal lobe and the right parahippocampal gyrus.

1.3 Non-numerical Covariates and Congruency Effect in Numerosity Comparison Tasks

Despite the advances made in ANS research, a significant debate remains regarding how such numerosity-tuned neural responses are derived from visual inputs. Given the inherent connection between numerosity and various visual set characteristics, such as the tendency for more items to occupy larger areas or be more densely spaced, determining the extent to which the neural response to numerosity differs from responses to other visual attributes is challenging.

In an attempt to minimize the influence of non-numerical continuous magnitudes, studies have employed various approaches while recognizing the impossibility of creating two sets of items that differ exclusively in numerosity (Gebuis et al., 2016; Gebuis & Reynvoet, 2012b). Considering studies utilizing dot-comparison tasks, a commonly used strategy involves structuring the arrays so that specific visual cues, including dot area (either cumulative surface area or average dot area, which are directly correlated), convex hull (the smallest contour around the dot array) or dot density (commonly calculated as cumulative surface area/convex hull), do not consistently align with numerical values across all trials (Gebuis et al., 2016; Gilmore et al., 2016; Wilkey & Ansari, 2020). To achieve this, stimuli are often designed such that, in half of the trials, the arrays are balanced in terms of extrinsic factors (e.g., cumulative surface area of the dots), while the intrinsic factors (such as dot diameter) are randomly altered. Conversely, in the remaining half of the trials, this procedure is reversed (Dehaene et al., 2005; Smets et al., 2015). This approach operates under the assumption that participants are discouraged from relying on non-numerical visual cues correlating with numerosity, given that no single visual cue reliably predicts numerosity across all task trials. However, studies have pointed out that participants may still shift focus between different visual properties across trials or integrate multiple visual cues on certain trials, meaning when one cue is uninformative, other one(s) might be used (Gebuis & Reynvoet, 2011, 2012b; Szűcs et al., 2013).

To address this concern, Gebuis and Reynvoet (2011, 2012a) devised an alternative algorithm capable of manipulating visual features of stimuli more accurately by controlling multiple visual cues to ensure that all visual parameters are uninformative about numerosity across all trials. Their method generates at the same numerical ratios both congruent trials, where the more numerous array exhibits one or more visual cues positively correlated with numerosity, and incongruent trials, where one or more visual cues negatively correlate with numerosity. Thus, during incongruent trials, participants are required to select the more numerous arrays solely based on discrete quantity, even when other visual cues may mislead the judgment (Wilkey & Ansari, 2020).

Nevertheless, even when non-numerical visual parameters are stringently controlled, research has shown that a congruency effect, which is the performance gap between congruent and incongruent trials, correlates with an index of the ANS, such as the Weber fraction (Szűcs et al., 2013). While the strength of this effect tends to diminish during typical development, it remains commonly observed in adulthood (Starr et al., 2017; Szűcs et al., 2013; Tokita & Ishiguchi, 2013). Generally, participants tend to be more accurate on the congruent trials (e.g., Cappelletti et al., 2014; Rodríguez & Ferreira, 2023; Szűcs et al., 2013), but a number of studies have found the reverse pattern; more accurate performance during incongruent than congruent ones (Gebuis & van der Smagt, 2011; Lee & Cho, 2019). Thus, even in instances where there is no correlation between numerosity and visual features across trials, visual features can still significantly influence numerosity processing on an individual trial basis. Consequently, task accuracy has been found to be better predicted by different combinations of visual cues than the difficulty of the numerosity ratio, casting doubt on the task's ability to effectively reflect a domain-specific ANS (Gebuis & Reynvoet, 2012a).

1.3.1 A Link between Numerosity Comparison Task and Domain-General Factors

The uncertainty brought by the congruency effect has prompted some scholars to suggest that there may not be a distinct cognitive mechanism exclusively dedicated to number processing, such as the ANS, but instead, humans integrate information from multiple non-numerical visual cues to inform numerosity (Gebuis & Reynvoet, 2012a; Leibovich et al., 2017). An alternative, more moderate view questions the suitability and validity of the numerosity comparison task itself. Several studies have proposed that performance on the task might depend on domain-general executive functions, enabling individuals to selectively attend to numerosity information while suppressing or controlling the influence of non-numerical visual parameters within the dot arrays (Formoso et al., 2017; Lee & Cho, 2019; Szűcs et al., 2013; Rodríguez & Ferreira, 2023). This latter notion finds support in studies indicating that the ability to discriminate numerosities correlates with components of executive function, such as VWM and inhibitory control.

A noteworthy study by Lee and Cho (2019) involving university undergraduate students demonstrated a connection between VWM capacity and the impact of numerical versus non-numerical visual parameters on participants' performance in a numerosity comparison task. Specifically, participants with lower scores in the VWM task displayed a more pronounced reverse congruency effect (i.e., a greater accuracy during incongruent trials than congruent ones) compared to those with higher scores in the VWM task. Additionally, no significant difference was found between low and high verbal working memory (henceforth, VbWM) groups regarding susceptibility to numerical versus non-numerical visual parameters. Further strengthening this finding, a recent study by Castaldi and colleagues (2021) demonstrated that adults' precision in numerosity comparison performance was bidirectionally influenced by a concurrent VWM task but not by a concurrent VbWM task. Additionally, participants showed an increased bias towards non-numerical visual cues during the visual, but not verbal, working memory task. As proposed by the latter study's authors, the findings of these studies indicate that numerosity judgment and the visual component of working memory may overlap in their resources.

Further insight into the relationship between numerosity comparison task performance and VWM has been provided by research in developmental dyscalculia (DD) and mathematics. For instance, Bugden and Ansari (2016) noted that the disparity in numerical acuity between individuals with DD and age-matched typically developing (TD) controls, as previously demonstrated, was significant only during the incongruent trials of their study, with no notable distinction during congruent trials. Moreover, the study found that individual differences in children with DD during incongruent trials were strongly predicted by their VWM performance. Similarly, other studies have highlighted that both in children with DD and their TD peers, the relationship between numerical acuity and mathematical achievement is significant only in the incongruent but not in the congruent trials (Fuhs & McNeil, 2013; Gilmore et al., 2013; Wilkey et al., 2020). Furthermore, when studies have entered VWM and intelligence (Coolen et al.,

2022) or inhibitory control (Fuhs & McNeil, 2013; Gilmore et al., 2013) as a predictor in a regression analysis, the relation between numerical acuity and mathematical achievement ceased to exist.

Taken together, multiple studies offer compelling evidence suggesting that performance in numerosity comparison task may not be solely indicative of the ANS but rather a composite of abilities. More specifically, these investigations posit that numerical acuity may be impacted by domain-general functions, particularly by the visual component of working memory.

1.2 Visual Working Memory

Working memory, a system of components that temporarily retains a limited amount of information in a heightened state of availability, forms an interface between memory, attention, and perception (Adams et al., 2018; Baddeley, 1998). It allows people to hold and manipulate relevant information in mind for ongoing processing tasks for a short period of time while simultaneously processing new information or executing cognitive tasks, such as reading, problem-solving, and learning (Baddeley, 2020; Baddley & Hitch, 1974; Cowan, 2008). The active involvement of working memory in various cognitive activities, as well as its role as a significant predictor of mathematical skills, has been underlined by numerous studies (e.g., Friso-Van den Bos et al., 2013; Passolunghi & Lanfranchi, 2012; Peng et al., 2016), positioning it as a central focus for a broad spectrum of research endeavors.

Distinguished from other forms of memory, working memory is recognized for its dual role in storing and processing information, whereas short-term memory (STM) has been primarily associated with the transient storage of sensory and cognitive information (Aben et al., 2012; Baddeley, 1986). Although consensus on the precise conceptualization of working memory remains elusive (Chai et al., 2018), the multicomponent model proposed by Baddeley and Hitch (1974; Baddeley, 1986, 2003) offers a widely accepted framework. According to this model, working memory consists of a domain-general central executive system of limited attentional capacity that controls and monitors information processing, along with two domain-specific subsystems for storing information: the phonological loop and the visuospatial sketchpad. The phonological loop is responsible for maintaining phonological information through controlled articulation processes, while the visuospatial sketchpad preserves spatial and visual information, facilitating the construction and manipulation of mental images. Moreover, these two domain-specific subsystems have been referred to as the STM of the model (Aben et al., 2012; Wang & Carr, 2014).

Contemporary research commonly interprets VWM and VbWM as outcomes of the interactions between the central executive and their storage systems – the visuospatial sketchpad and the phonological loop, respectively (Allen et al., 2017; De Beni et al., 2005; Wang & Carr, 2014). These definitions of VWM and VbWM are considered theoretically more economical than definitions positing the presence of domain-specific active processors functioning independently on visuospatial and verbal STMs (Wang & Carr, 2014). Thus, in alignment with contemporary research, this thesis also adopts the conceptualization of VWM as a result of the interaction between the central executive and its dedicated storage system, handling the temporary retention and manipulation of visual and spatial information during real-time tasks.

1.2.1 Capacity Limits and Filtering Efficiency

The ability to actively hold visual information is widely acknowledged as significantly limited, with adults typically able to maintain a maximum of three to four items concurrently within VWM (Cowan, 2001; Luck & Vogel, 1997; Pashler, 1988). The capacity limit undergoes developmental changes during childhood, with some studies indicating a mature level reached in adolescence (Riggs et al., 2006; Van Leijenhorst et al., 2007), while others suggest it may not stabilize until young adulthood around the age of 20, after which a gradual decline with age is observed (Brockmole & Logie, 2013; Luna et al., 2004). Moreover, individuals display considerable variability in their VWM capacity levels, which being significantly linked to various cognitive functions (Fukuda et al., 2010; Johnson et al., 2013; Unsworth et al., 2015; Vogel et al., 2001; Vogel & Machizawa, 2004), makes it a valuable tool for investigating individual differences.

Although the exact underlying mechanism of VWM capacity limits remains a matter of ongoing debate, it is commonly assumed that estimates of VWM capacity encapsulate the full range of the system's capabilities (Alvarez & Cavanagh, 2004; Bengson & Luck, 2016; Luck & Vogel, 2013; Zhang & Luck, 2008). However, recent research has illuminated that this capacity is not solely contingent upon the number of items or features that can potentially be stored but also on the efficiency of item storage. For example, a number of studies have demonstrated that participants who could recall more objects from spatial arrays also excluded irrelevant salient objects more efficiently and vice versa (Cowan & Morey, 2006; Gaspar et al., 2016; Vogel et al., 2005). Therefore, a crucial role in the regulation of access to VWM storage is thought to be played by selective attentional processes, responsible for prioritizing relevant information while filtering out irrelevant data (Awh et al., 2006; Cowan & Morey, 2006; McNab & Klingberg, 2008; Plebanek & Sloutsky, 2019; Vogel et al., 2005). Even though attentional control was already assigned an important role in Baddeley's working memory model, the recent findings provide more support for the close relationship between working memory and attention by suggesting that the efficacy of information filtering contributes significantly to an individual's overall VWM capacity level.

The filtering-efficiency hypothesis specifically proposes that attention regulates the influx of sensory data into the limited-capacity VWM system and that including irrelevant information reduces storage capacity for task-relevant items (Gaspar et al., 2016). It provides an explanation why individuals with lower VWM capacity may, in certain situations, retain more information in memory than those with higher VWM capacity: high-capacity individuals encode only task-relevant items, whereas low-capacity individuals encode both irrelevant and relevant items (Vogel et al., 2005). This hypothesis has received considerable support (Robison et al., 2018), such as studies demonstrating that targeted training on filtering efficiency can enhance individuals' VWM capacity (Li et al., 2017; Schmicker et al., 2016) or that age-dependent decline in VWM capacity is coupled with reduced filtering scores (Jost et al., 2011).

1.2.2 Neural Substrate of Visual Working Memory

Research exploring the neural mechanisms underlying VWM capacity has utilized fMRI and electroencephalography techniques, revealing a region within the posterior parietal cortex (PPC), including the IPS, where neural activity reflects the quantity of stored visuospatial information (Linden et al., 2003; Todd & Marois, 2004; Vogel & Machizawa, 2004). This parietal involvement in VWM maintenance has received support from neuropsychological lesion studies showing that patients with damage to the right PPC struggle to remember a small number of sequentially presented objects and locations over short delays (Berryhill & Olson, 2008). Additionally, prefrontal cortical regions have been postulated to play a pivotal role in modulating attentional control over access to VWM storage in the parietal cortex (Curtis & D'Esposito, 2003; McNab & Klingberg, 2008). However, despite a consensus emerging regarding VWM capacity being mediated by a network of frontoparietal and sensory cortical areas (Darki & Klingberg, 2015; Klingberg et al., 2002; Owen et al., 2005; Rottschy et al., 2012), an active debate persists regarding the specific functional role(s) of these brain regions (Ester et al., 2015).

1.2.3 Measuring Visual Working Memory Capacity

A commonly used method to estimate individuals' VWM capacity involves a change detection task, initially proposed by Phillips in 1974 and later popularized by Luck and Vogel in 1997. This paradigm features two well-known versions: a single-probed and whole-display recognition task (Barton & Brewer, 2013; Rouder et al., 2011). In both versions, participants are presented with a memory array of visually presented objects (usually for a period between 100 and 500ms), followed by a brief delay (generally about 1000ms) before a probe array appears. In the single-probed recognition, only one target, either a studied or a novel item, is presented at one of the studied locations on the probe array. Alternatively, in the whole-display recognition, the probe array is either identical to the memory array or has replaced one of its items with a novel item. In both versions of the task, participants must provide an un-speeded two-alternative forced-choice response on each trial, indicating whether or not a change was detected.

To index individuals' VWM capacity in the whole-display task, Pashler (1988) introduced a measure, denoted by $\hat{k}_p$, which remains widely utilized today:

$$\hat{k}_p = N\left(\frac{\hat{h} - \hat{f}}{1 - \hat{f}}\right). \tag{1}$$

In this equation, *N* indicates the number of items in the sample array, $\hat{h}$ the observed hit rates, and $\hat{f}$ the false alarm rates on no-change trials. The measure is based on the idea that on each trial, *k* out of *N* items can be held in memory, and if a changed item coincides with the one stored in working memory (with a probability of *k*/*N*), then the changed item will be detected. Alternatively, for the single-probed task, a slightly modified measure, denoted by $\hat{k}_c$, was introduced by Cowan (2001) to account for the assumption that if the cued item is in WM, the correct answer will be available whether or not there is a change in the test array:

$$\hat{k}_c = N(\hat{h} - \hat{f}). \tag{2}$$

This is assumed to happen in *k*/*N* of the trials where a change is present and (in contrast to Pashler, 1988) in *k*/*N* of the trials where there is no change (Cowan et al., 2006).

Even though the two equations were developed for two different versions of the change detection task, both are grounded in the same psychological model: both adhere to the concept of a fixed capacity and all-or-none encoding within working memory (Morey, 2011). Some studies have interchangeably applied these equations across the two versions of the task, and some even reported both measures within the same dataset (e.g., Lee et al., 2010; Treisman & Zhang, 2006; Vogel et al., 2005). However, Rouder and colleagues (2011) strongly advocate for selecting $\hat{k}_p$ or $\hat{k}_c$ based on the specific experimental design, as this choice could significantly impact the evaluation of how VWM capacity correlates with other variables.

1.3 Present Study

The objective of this thesis is to explore the relationship between a frequently used paradigm assessing ANS ability, the numerosity comparison task, and a domain-general function, specifically the visual component of working memory. Traditionally, the ANS has been understood as a distinct non-verbal cognitive system facilitating the estimation, comparison, and basic numerical operations of non-symbolic numerosities. However, emerging research suggests that performance in numerosity comparison tasks may not exclusively assess the domain-specific ANS; rather, it might also involve the utilization of abilities associated with domain-general executive function, particularly VWM. Consequently, this task may not solely reflect the precision of internal quantity representations, but instead, a combination of abilities.

The inherent challenge of the numerosity comparison task lies in constructing two sets of items, such as arrays of dots, that do not intertwine with non-numerical visual factors that are closely associated with numerosity perception. The importance of these non-numerical visual covariates has been underscored by recent research, suggesting that individual differences in the VWM capacity may significantly affect the susceptibility to numerical versus non-numerical visual information. Specifically, individuals with higher VWM capacity may be better able to selectively attend to numerosity information while effectively suppressing the influence of non-numerical visual parameters, resulting in better numerosity discrimination accuracy. Further supporting a link between these two functions, findings from neuroimaging have provided evidence that both numerosity processing and VWM performance activate areas on the frontoparietal network, particularly the IPS of the PPC. Numerical cognition studies have revealed that IPS contains numerosity-tuned neuronal populations, and conversely, studies investigating working memory have found that the activity in IPS can reflect the quantity of stored visuospatial information.

Overall, the evidence outlined in this literature review suggests that VWM may play a crucial role in the performance of numerosity comparison tasks. Specifically, studies suggest that VWM may enable individuals to selectively attend to quantity while filtering out non-numerical distractions. However, this notion of a filtering mechanism, intertwined with non-numerical variables and influencing numerosity discrimination accuracy, is not new. A study by Piazza and colleagues (2018) has proposed that the increase in numerical acuity during development arises from the ability to focus on the relevant dimension of number while avoiding interference from irrelevant but often co-varying quantitative dimensions. However, the authors did not delineate a specific cognitive mechanism responsible for this filtering process, thus leaving an open question about whether it is domain-specific or domain-general function.

Therefore, the present thesis sets out to explore whether a functional relationship exists between numerical acuity and domain-general VWM, even when the non-numerical visual parameters are stringently controlled. Specifically, it aims to investigate whether there exists a correlation between the effect of non-numerical visual cues on numerosity discrimination accuracy and VWM capacity. To achieve this objective, this thesis will administer both a prototypical paradigm measuring VWM capacity, namely the whole-display version of the change detection task, and a numerosity comparison task with dot arrays generated by a recently published Matlab toolbox CUSTOM (De Marco & Cutini, 2020) algorithm.

The advantage of the CUSTOM algorithm is that it offers robust control over non-numerical covariates through various manipulation methods. In particular, the algorithm equips researchers to manipulate the size and area of dots via diameter, total surface area, or total contour (i.e., the sum of the perimeters of all the elements) while maintaining a desired convex hull value constant throughout the task. Additionally, when opting to maintain either total surface area or total contour constant, the program allows for random variability in dot diameters within a controlled range, resulting in dots of varying sizes. The algorithm also puts special consideration over density. Given that density is traditionally calculated as the total surface divided by the convex hull, the algorithm ensures stringent regulation for this metric when balancing the total dot surface area. Additionally, it introduces another density index, the average distance between dots (henceforth, ADBD), which can be assessed as a posteriori array creation.

Notably, the developers of the CUSTOM algorithm demonstrated in their study that when the total surface area remains constant, the total contour exhibits a positive correlation with numerosity, while ADBD decreases with increasing numerosity. Conversely, when total contour is kept constant, total surface and classic density show a negative correlation with numerosity, while ADBD increases until reaching a plateau for numerosities above 30. Nonetheless, maintaining a constant total surface area or total contour results in a reduction in the average diameter of dots as numerosity increases. This implies that under conditions where total surface area or total contour is balanced, dot arrays will consistently contain visual variables correlating negatively and positively with numerosity. Overall, the algorithm efficiently generates high-precision visual stimuli within an impartial theoretical framework, offering robust control over visual covariates, which is crucial for numerosity comparison tasks.

The present thesis will create a numerosity comparison task comprising eight numerosity ratio conditions of varying difficulty. Half of the task trials will be equated based on the total contour of the dots, while the other half will be equated based on the total dot surface area. Consequently, both types of equated trials will exhibit partial (in)congruence, incorporating visual variables that positively and negatively correlate with numerosity in a manner described earlier. This thesis will then investigate whether overall task accuracy, aggregated over employed numerosity ratios of the numerosity comparison task, or the difference in accuracy between the two equated trial types correlates with VWM capacity. Accuracy will be utilized as an index of numerical acuity as research has demonstrated it to be the optimal measure when studies adopt a correlational standpoint.

Given the rigorous control over visual covariates by the CUSTOM algorithm and the fact that prior studies have mainly shown a link between a congruency effect or incongruent trials and domain-general factors, it is hypothesized that numerical acuity will not directly correlate with VWM capacity. However, despite the strict controls, the algorithm does not entirely remove the influence of non-numerical visual covariates, as the two types of equated trials exhibit different non-numerical visual cues covarying with numerosity. Therefore, it is hypothesized that a significant difference in accuracy will be found between the two equated trial types. Furthermore, it is expected that this difference in accuracy will correlate with VWM capacity.

2 Methods

2.1 Participants

A total of 46 students (19 male, 27 female) from the local University aged 19 to 34 years (M = 22.70, SD = 2.68) took part in the present experiment. All had normal or corrected-to-normal vision and none reported a history of neurological or psychiatric disorders. After the nature of the study was explained, each participant gave their written informed consent. The research protocol was approved by the local Ethical Committee (Protocol #4683), and the examinations were conducted in accordance with the tenets of the Declaration of Helsinki.

2.2 Experimental Procedure and Stimuli

The experiment took place in a softly lit chamber within the research facilities of the General Psychology department at the University of Padova. Participants were comfortably seated facing a desk outfitted with a computer screen and a standard keyboard. Stimuli for the change detection and numerosity comparison tasks were generated using E-Prime 2 software (Psychology Software Tools Inc.) and Matlab (Version: 9.13.0, R2022b, The MathWorks Inc.), respectively. Both tasks were presented on a black background (RGB: 0, 0, 0) of a 24” CRT monitor with a refresh rate of 60Hz positioned at a visual distance of approximately 65cm. The tasks were displayed sequentially with a short break in between, and the order of presentation was counterbalanced across participants.

2.2.1 Numerosity Comparison Task

Each trial of the numerosity comparison task started with a 600ms display of a light-grey (RGB: 220, 220, 220) central fixation cross (0.8° x 0.8°) followed by a presentation of two gray squares (of 3° x 3° each and a square-center to screen-center distance of 2.5°) positioned on the left and right side of the fixation cross for a randomized duration between 1500 to1700ms. After this, one of the two gray squares served as a reference array, displaying either 16 or 32 white dots, and the other gray square, i.e., paired array, had a systematically varied number of dots according to the reference array. For the 16-dot reference array, the paired array contained 12, 13, 14, 15, 17, 18, 19, or 20 dots, and for the 32-dot reference, the paired array had 24, 26, 28, 30, 34, 36, 38 or 40 dots. The side of the reference array during the task was randomized. Overall, eight distinct numerical (small/large) ratio conditions were created, including 0.750, 0.813, 0.842, 0.875, 0.800, 0.889, 0.938, and 0.941.

A Matlab toolbox CUSTOM (De Marco & Cutini, 2020) algorithm was used to generate the task arrays. This ensured that in 50% of the trials, the dots on the reference and paired array had equal total contour, and in the remaining 50% of trials, the dots covered an equal total surface area on both arrays. Furthermore, throughout the task, the convex hull value of the arrays was kept constant. The task compromised ten practice trials and a total of 192 experimental trials, which were randomly selected from a set of 1280 pairs of arrays of dots generated for each participant. The experimental trials were divided into 12 blocks of 16 trials each.

Prior starting the task, the participants were given the instruction to maintain their gaze at the fixation cross and indicate the array with a larger numerosity of dots by pressing a corresponding key on a keyboard ("A" for the left and "L" for the right-sided array). The dot arrays remained displayed until the participants gave their response within a maximum response interval of 2000ms. A visual illustration of the stimuli and the sequence of events of the numerosity comparison task is presented in Figure 1.

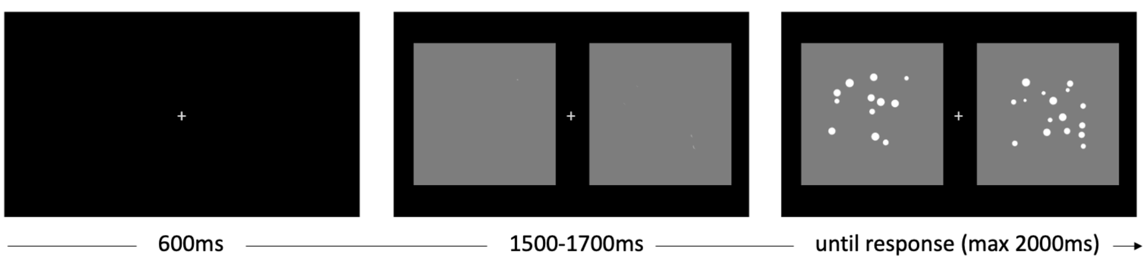

Figure 1. A schematic illustration of the sequence of events on one trial of the numerosity comparison task. Participants were instructed to maintain their gaze at the fixation cross and indicate whether the left or right array had more white dots by pressing a corresponding key on a computer keyboard. In the present example trial, the 16-dot reference array (right) and the 12-dot paired array (left) cover an equal total dot surface area.

2.2.2 Change Detection Task

The change detection task consisted of 16 practice trials followed by a total of 810 experimental trials, which were organized into 27 blocks of 30 trials each. Each trial commenced with a participant pressing a spacebar on the computer keyboard, initiating the display of a light gray (RGB: 220, 220, 220) central fixation cross (0.8×0.8°) that jittered randomly in steps of 20ms for a duration between 900 and 1000ms. Subsequently, a memory array was presented for 150ms holding a notional 5.5° x 5.5° rectangular region and featuring either two, three, or five colored squares around the fixation cross. The squares (each having a size of 1°×1°) were placed at random positions of the memory array but constrained with a minimum distance of 1.5° between their upper left corners of two adjacent squares and 1.3° between the fixation cross and the side of the nearest square. Colors for the squares were selected randomly from a pool of 11 colors, including red (RGB: 255, 0, 0), light green (RGB: 0,255,0), blue (RGB: 0, 0, 255), yellow (RGB: 255, 255, 0), magenta (RGB: 255, 0, 255), cyan (RGB: 0, 255, 255), dark green (RGB: 30, 140, 60), purple (RGB: 128, 0, 255), orange (RGB: 255, 128, 0), brown (RGB: 157, 0, 23), and pink (RGB: 255, 174, 201).

Participants were instructed to maintain their gaze at the fixation cross and memorize both the color and location of all squares displayed on the memory array. Following a 900ms blank retention interval, a probe array was presented that was either identical to the content of the memory array or had changed one of its square item's colors to a novel one, with an equal likelihood. The probe array remained visible until the participants gave their response by pressing a corresponding key on the keyboard to indicate whether or not they had detected a change (i.e., "1" for the item is changed and "2" for no change, keys counterbalanced across participants). A schematic illustration detailing the stimuli and sequence of events in the change detection task is provided in Figure 2.

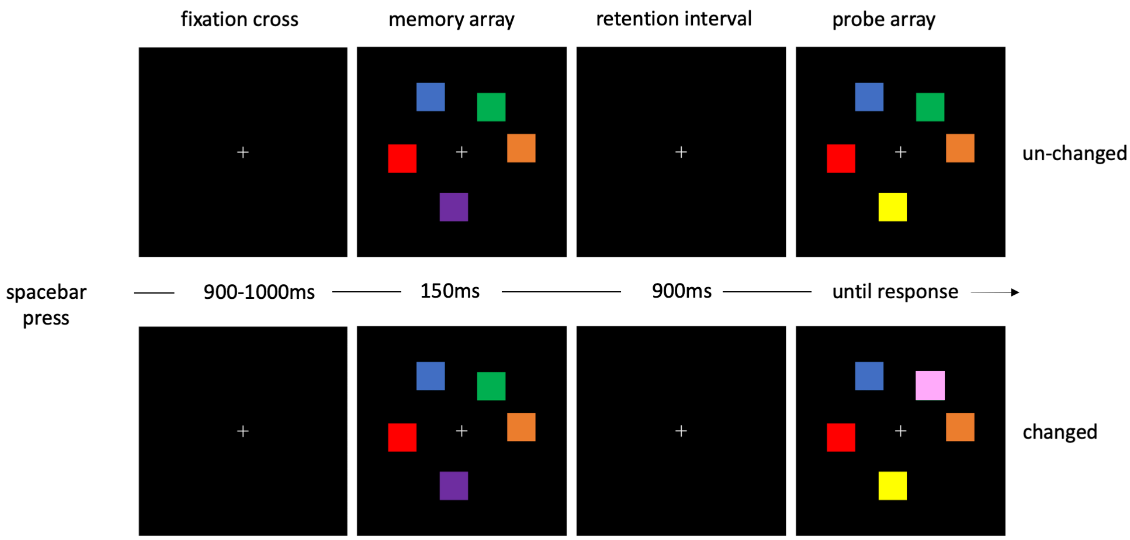

Figure 2. A schematic illustration of the sequence of events on the change detection task. Participants were instructed to maintain their gaze at the fixation cross and to memorize the color and location of the squares from the memory array. After the retention interval, participants were asked to indicate whether the probe array had remained identical (un-changed trial example, upper row) or changed one of the memory array's target colors (changed trial example, lower row). The present un-changed and changed trial examples show a set size of five items.

2.3 Data Analysis

Research indicates that valid reaction times typically commence after 100-200ms, allowing for essential physiological processes, such as stimulus perception and motor responses (Luce, 1991; Whelan, 2008). Therefore, consistent with Miller's (2023) approach, trials with response times under 150ms were excluded from the numerosity comparison dataset, representing less than 0.6% of the total task trials. Similarly, for the change detection task, trials with response times below 150ms were eliminated as outliers, as well as those surpassing five seconds in order to align with methodologies observed in prior change-detection studies (e.g., Gilchrist & Cowan, 2014; Udale et al., 2018). This resulted in the removal of less than 1.7% of the task trials. Furthermore, four participants exhibiting below-chance accuracy (<0.5) across the three different set size conditions of the change detection task were excluded from the analyses. The data analyses were conducted using Microsoft Excel and the statistical software program SPSS, with an alpha level set at $\alpha < 0.05$ to determine statistical significance.

3 Results

3.1 Numerical Acuity

The total accuracy proportions averaged across the employed eight different numerosity ratios were used as an index to determine each participant's numerical acuity. On average, participants exhibited an accuracy proportion of 0.76 (*SD* = 0.08), indicating a generally high level of proficiency in correctly identifying the more numerous array than compared to a chance performance level. Furthermore, consistent with prior research, the average accuracy proportions demonstrated a gradual decline as the numerosity ratio approached 1. This trend of increasing difficulty in discriminating numerosities accurately as the numerosity ratio increases is illustrated in Figure 3.

To explore potential variations in accuracy between trials featuring a reference frame containing 32 dots versus 16 dots (with accuracy averaged across employed numerosity ratio conditions), a paired-sample t-test was conducted. Examination of a boxplot revealed no outliers, and the difference in accuracy between the two reference frame trials demonstrated a normal distribution, as confirmed by Shapiro-Wilk's test ($p = .850$). The average proportion of correct responses did not show a statistically significant difference between the trials with 32-dot (M = 0.77, SD = 0.09) and 16-dot reference frames (M = 0.76, SD = 0.08), $t(45) = 1.68$, $p = .099$. Therefore, manipulating the number of dots in the reference frame did not significantly affect participants' performance on the numerosity comparison task.

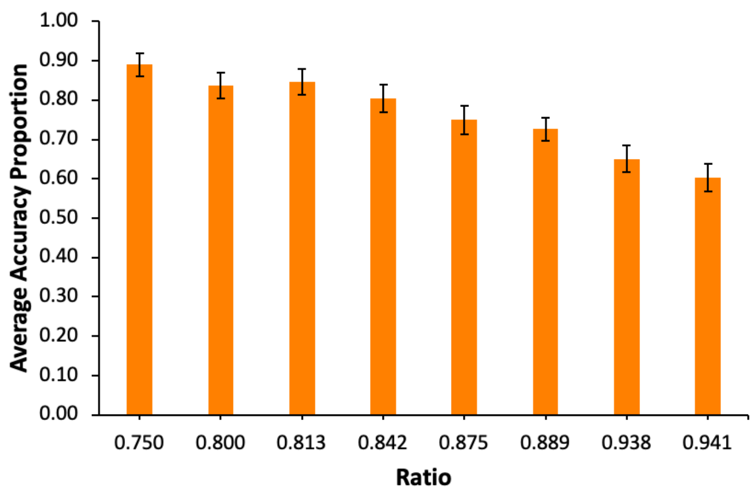

Figure 3. Participants' average accuracy scores are presented as a function of the numerical ratio between the two dot arrays in the numerosity comparison task. Error bars represent between-subject confidence intervals (95%, two-tailed).

Similarly, to determine whether the proportion of correct responses varied between trials with equated total contour and total dot surface area (with average accuracy aggregated across employed numerosity ratio conditions), a paired-sample t-test was employed. The difference in accuracy proportion between the two types of equated trials demonstrated a normal distribution, as assessed by Shapiro-Wilk's test ($p = .113$). Although one outlier was observed in boxplot inspection, subsequent analyses with and without the outlier remained consistent, and thus, it was retained in the analysis. Accuracy proportions were found to be higher on trials with the total dot surface area equated (*M* = 0.77, *SD* = 0.08) compared to trials with total contour equated (*M* = 0.75, *SD* = 0.09), showing a statistically significant 0.03 mean increase in the proportion of correct responses, 95% CI [0.01, 0.04], $t(45) = 2.81$, $p = .007$, with a small effect size, $d = .42$. This indicates that performance on the numerosity comparison task was significantly influenced by non-numerical visual parameters, thereby compromising the internal consistency of the task.

3.2 Visual Working Memory Capacity

The potential differences in participants' average accuracy proportions across the three set size conditions were investigated by performing a one-way repeated measures ANOVA. Three outliers were identified via boxplot examination, and violations of normal distribution were noted for set sizes two and three, as indicated by the Shapiro-Wilk test ($p < .05$). However, subsequent analysis, with the outliers or without the outliers and having the normality assumption met for all set sizes, did not yield in different conclusions; therefore, the outliers were retained in the analysis. The average accuracy proportions exhibited statistically significant differences across set sizes, $F(2, 82) = 246.87$, $p < .001$, partial $\eta^2 = .86$, with accuracy progressively decreasing from set size two (*M* = 0.94, *SD* = 0.04) to set size three (*M* = 0.91, *SD* = 0.05) and further to set size five (*M* = 0.80, *SD* = 0.07). Post hoc comparison using a Bonferroni adjustment revealed that the decreases in accuracy were statistically significant between all set sizes, indicating heightened task difficulty as the number of squares to remember increased from set size two to set size three (*M* = 0.04, 95% CI [0.03, 0.05], $p < .001$), from set size two to set size five (*M* = 0.15, 95% CI [0.12, 0.17], $p < .001$), and from set size three to set size five (*M* = 0.11, 95% CI [0.09, 0.12], $p < .001$).

To determine the average number of colored square items each participant memorized in the change detection task, *K*, derived from Pashler's (1988) Equation (1), was used as an index. Upon examining the *K* values between different set sizes, two outliers were detected through boxplot analysis, along with a violation of normality in all set sizes, as indicated by the Shapiro-Wilk test ($p < .05$). Therefore, to assess potential variations in participants' *K* values between the set sizes, Friedman test was employed due to its robustness to normality violations. The participants' *K* values showed statistically significant changes across set sizes, $\chi^2(2) = 84.000$, $p < .001$, and post hoc pairwise comparisons performed with a Bonferroni correction revealed the differences to be statistically significant between all conditions. Specifically, *K* values significantly increased from set size two (*Mdn* = 1.92) to set size three (*Mdn* = 2.79) ($p < .001$), from set size two to set size five (*Mdn* = 4.43) ($p < .001$), and from set size three to set size five ($p < .001$). This means that VWM capacity, between these set size conditions, reached its maximum level at set size five, as illustrated in Figure 4.

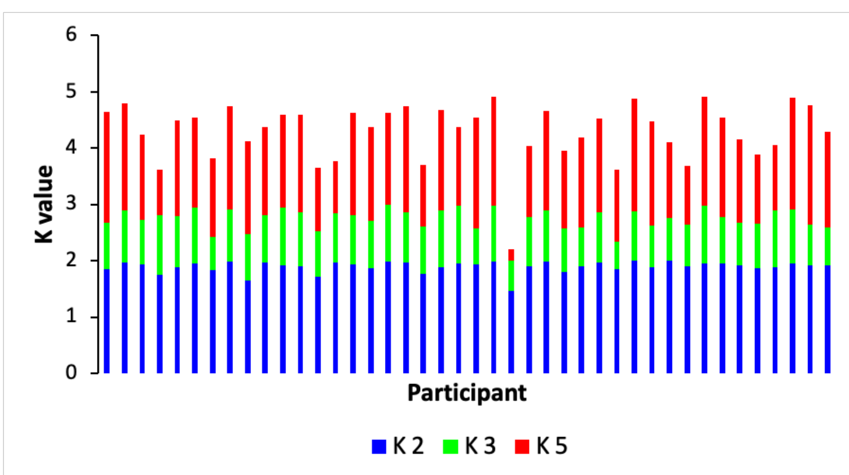

Figure 4. Participants' (n = 42) K values for each set size in the change detection task reported as cumulative sum column.

To assess the reliability of the *K* values at set size five, an even-odd split-half analysis using Spearman's rank-order correlation was conducted. Visual examination of a scatterplot confirmed the relationship to be monotonic, although one outlier was identified. As the removal of the outlier did not yield an appreciable difference in the results, it was kept in the analysis. The analysis showed a statistically significant positive correlation in *K* values at set size five between even and odd trials, $r_s(40) = .62$, $p < .001$, indicating strong internal reliability of the change detection task measuring *K* values (Cohen, 1988).

Additionally, d' value (Green & Swets, 1966) was employed to gauge the strength of VWM item representations. Specifically, the d' can be used to assess participant's ability to retain and accurately recall visual information from VWM over a brief duration. The formula used to calculate d' was:

$$d' = Z(\hat{h}) - Z(\hat{f})$$

where $Z(\hat{h})$ denotes the Z-score of the hit rate and $Z(\hat{f})$ the Z-score of the false alarm rate. To explore potential variations in d' values across set sizes, a one-way repeated measures ANOVA was conducted. Although three outliers were identified, the data was normally distributed at each set size, confirmed by the boxplot and Shapiro-Wilk test (p > .05), respectively. Given that subsequent analyses yielded consistent results regardless of the outliers' inclusion, they were retained in the analysis. Participants' d' values exhibited significant variation across set sizes, $F(2, 82) = 182.98$, $p < .001$, partial $\eta^2 = .82$, with d' values progressively declining from set size two ($M = 3.38$, $SD = 0.68$) to set size three ($M = 2.81$, $SD = 0.62$), and further to set size five ($M = 1.96$, $SD = 0.62$). Post hoc comparison with Bonferroni correction revealed statistically significant decreases in *d′* value from set size two to set size three (M = 0.57, 95% CI [0.39, 0.75], $p < .001$), from set size two to set size five (M = 1.42, 95% CI [1.21, 1.63], $p < .001$), and from set size three to set size five (M = 0.85, 95% CI [0.69, 1.02], $p < .001$), indicating a progressive decline in working memory signal strength as the number of items to be remembered increased across the set sizes.

3.3 Correlation Between Numerical Acuity and Visual Working Memory Capacity

A Spearman's rank-order correlation was employed to explore a potential relationship between participants' numerical acuity (numerosity discrimination accuracy across employed ratios) and their VWM capacity (maximum K value attained at set size five). Initial data inspection suggested an approximately monotonic relationship, albeit with the presence of one outlier. As the removal of this outlier did not result in significant alterations in the outcome, its inclusion was validated in the analysis. Nevertheless, the findings unveiled no statistically significant correlation between numerical acuity and VWM capacity, $r_s(40) = .18$, $p = .257$, suggesting no direct link between these performance metrics. For visual clarity, Figure 5a illustrates a scatterplot depicting the relationship between these values.

Finally, to assess whether there was any association between participants' VWM capacity and the difference in numerical acuity between the two equated trial types of the numerosity comparison task (numerosity discrimination accuracy in both equated trial types averaged across employed ratios), a Spearman rank-order correlation was conducted. Although the scatterplot suggested an approximately monotonic relationship, it contained two outliers (Figure 5b). Nonetheless, subsequent analyses, whether excluding or including the outliers, yielded consistent results, prompting their retention in the analysis. However, a non-significant relationship with a very small effect size emerged between the difference in numerosity discrimination accuracy proportions between the two equated trial types and *K* values at set size five, $r_s(40) = -.064$, $p = .686$. This indicates that VWM capacity did not exert a significant influence on the extent to which the non-numerical visual parameters impacted participants' performance accuracy in the numerosity comparison task. Notably, a power analysis (using G*Power 3.1 software) indicated that the achieved statistical power of 0.07 was low for detecting the observed Spearman correlation coefficient, given the used sample size (n = 42). To achieve a statistical power of 0.80 using alpha level of 0.05 and two-tailed test, a sample size of 1913 would have been needed for the observed Spearman correlation coefficient.

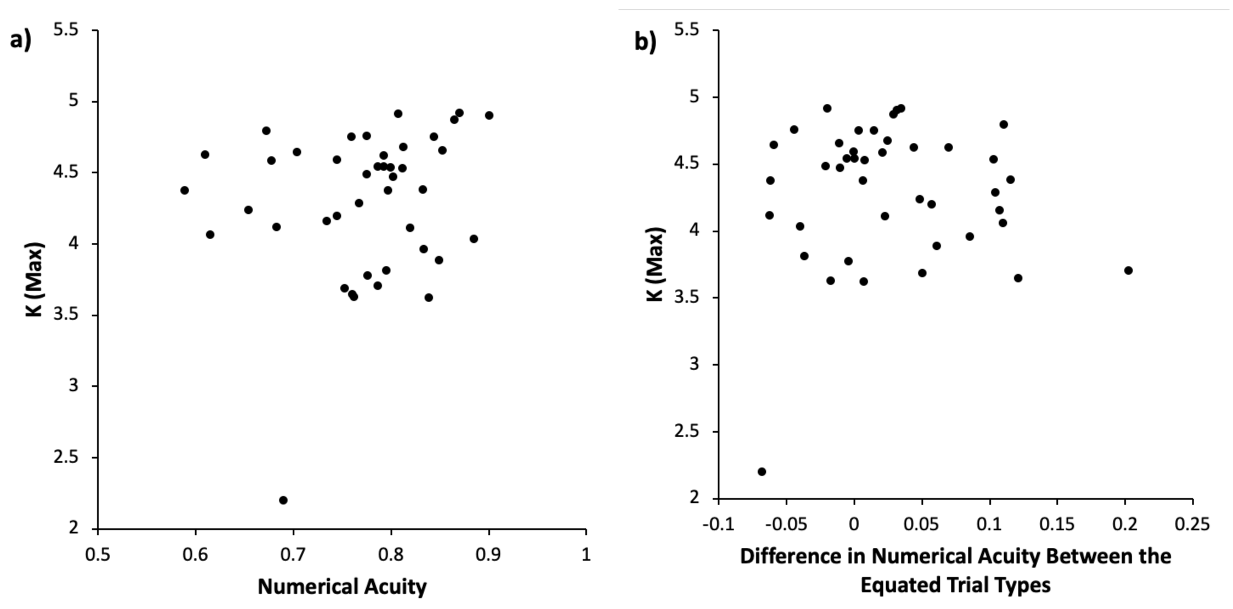

Figure 5. The figure presents scatterplots depicting (a) the association between numerical acuity (i.e., numerosity discrimination accuracy across employed numerosity ratios) in the numerosity comparison task and maximum *K* values in the change detection task and (b) the association between the difference in numerical acuity between trials with equated total dot surface and total contour in the numerosity comparison task and the maximum *K* values.

4 Discussion

The present thesis investigated whether the performance in numerosity comparison tasks is functionally connected to the visual component of working memory. Specifically, it explored whether non-numerical visual cues in two simultaneously presented dot arrays influenced the precision of numerosity discrimination, known as numerical acuity, and whether this effect was tied to the limited capacity of VWM. Confirming previous research, the results of the empirical investigation reported in the present thesis showed that discriminating numerosities becomes increasingly challenging as the ratio between the quantities of dots being compared approaches 1. In order to minimize the influence of non-numerical visual cues on numerical acuity, two highly equated trial types were generated for the numerosity comparison task via Matlab toolbox CUSTOM algorithm. Despite the meticulous visual controls, a significant difference in numerical acuity was found between the two employed equated trial types. Specifically, numerosity discrimination accuracy was higher on trials when the total dot surface area remained constant compared to when the total contour of the dots was held constant. Therefore, the findings strongly indicate that even when the best and most rigorous available controls are administered to the generation process of the numerosity comparison task, non-numerical visual cues seriously compromise the reliability of the task. This strongly suggests that the ability to discriminate numerosities relies not only on numerical factors but also on non-numerical visual information, which aligns with findings from previous research (e.g., Gebuis et al., 2016; Gebuis & Reynvoet, 2012a, 2012b; Leibovich et al., 2017; Szűcs et al., 2013).

The central focus of this thesis was to explore the connection between numerical acuity and VWM capacity. As expected, no direct inter-individual correlation was found between the numerical acuity and VWM capacity. This outcome is in line with previous research, which has primarily highlighted a relationship between domain-general executive functions and the congruency effect or incongruent trials of numerosity comparison tasks rather than directly with numerical acuity (e.g., Bugden & Ansari, 2016; Lee & Cho, 2019). However, surprisingly, even though a significant difference in accuracy was found between the rigorously equated trial types in the numerosity comparison task, the relationship between this disparity and VWM capacity was found to be non-significant with a very small effect size. In other words, individuals' level of VWM capacity did not significantly predict the extent to which non-numerical visual cues affected numerosity discrimination accuracy. Overall, this implies that while non-numerical visual covariates undoubtedly play a role in influencing our ability to discriminate numerosities, this ability likely operates on a distinct processing stage of VWM within human cognition.

Nevertheless, it is imperative to consider possible reasons for this unexpected null correlation between VWM capacity and the effect of non-numerical visual parameters on numerical acuity. Importantly, the results indicated that the present study may have suffered from a lack of statistical power. If the true effect size between these functions in the population is very small, as indicated by the findings, a larger sample size would have been needed to detect a statistically significant relationship. However, given the small detected effect size, increasing the sample size to achieve statistical significance might not provide substantial ecological value. Therefore, the current sample size could be considered sufficient.

An alternative option to consider for the absence of correlation could be that the study was unable to detect the true effect size due to the manner in which the VWM capacity index was obtained. While the *K* values exhibited a progressive increase with set size, concomitant with a decrease in *d′* values, indicating a diminishing ability to reliably differentiate between remembered and non-remembered items, it is worth noting that the *K* values at set size five may not have reached the maximum capacity of VWM for all participants. Hence, it is plausible that the employed change detection task may not have been optimal in detecting a memory plateau at the largest set size. Even though most studies report that adults can typically maintain a maximum of three to four items concurrently within VWM (e.g., Cowan, 2001; Luck & Vogel, 1997; Pashler, 1988) and using set sizes above individual's capacity limit (e.g., set size with eight items) can actually lead to significantly lower capacity estimate than using set size near-capacity (Fukuda et al., 2015), incorporating a fourth set size, such as six items, might have validly confirmed that each participant's VWM capacity reached its maximum.

Another limitation of the chosen method for assessing VWM capacity is that the *K* value was derived from a prototypical change detection task that did not include distractors. Consequently, participants were not required to focus solely on targets while filtering out distractors. Previous studies have specifically suggested that domain-general functions may enable individuals to selectively attend to quantity while filtering out non-numerical distractions (e.g., Formoso et al., 2017; Lee & Cho, 2019; Szűcs et al., 2013; Rodríguez & Ferreira, 2023). Therefore, merely assessing the VWM capacity without considering the filtering efficiency of VWM may not adequately demonstrate a significant relationship between VWM capacity and numerical acuity. Although previous research has illustrated an intricate link between attention-based filtering efficiency and VWM capacity (e.g., Awh et al., 2006; Cowan & Morey, 2006; Vogel et al., 2005), therefore, it could be assumed that this relationship is already reflected in simple capacity estimates. However, without directly accounting for the variance introduced by filtering efficiency, the full extent of VWM capacity may not be captured.

Previous studies specifically looking into the filtering process of VWM have utilized a modified version of the change detection task that incorporates trials with and without distractors, each containing the same total number of targets. This altered version enables the assessment of filtering cost, often used as an index of filtering efficiency (Allon & Luria, 2017; Vogel et al., 2005). The filtering cost can be calculated as the difference in accuracy between arrays containing only targets (e.g., three targets) and arrays with the same number of targets but also including distractors (e.g., three targets and three distractors). The derived score indicates the extent to which performance was impacted due to the introduction of distractors, i.e., a lower score signifies better filtering efficiency (Allon & Luria, 2017; Hadar et al., 2019).

Alternatively, neurophysiological measures offer another means to investigate VWM's filtering efficiency. Recent studies employing electroencephalography (EEG) have identified a lateralized event-related potential component in the EEG signal, referred to as the Contralateral Delay Activity (CDA). The CDA serves as a neural marker for the number of items encoded and sustained in VWM (Adam et al., 2018; Feldmann-Wüstefeld, 2021; Vogel & Machizawa, 2004) and its key feature lies in its amplitude increasing in correspondence to the number of objects held in VWM, leveling off asymptotically for object arrays meeting or surpassing an individual's storage capacity level (Vogel & Machizawa, 2004). The measurement of the CDA necessitates a lateralized version of the change detection task (or its close variant), which incorporates a cue before the onset of the memory array to indicate the relevant side for the current trial (Feldmann-Wüstefeld, 2021). This means that instead of attempting to recall all target items from both the right and left side of the memory array, participants focus solely on encoding and retaining target items from the cued side, as only items from that designated side will be probed at the end of the trial. Importantly, the memory arrays exhibit variability in the total number of items across trials, yet they maintain a balance between the right and left sides of the visual field. During the retention/delay period, the CDA presents sustained negativity over posterior electrode sites in the contralateral electrodes relative to ipsilateral electrodes (Vogel & Machizawa, 2004). Thus, the CDA involves measuring the difference in amplitude between the contralateral and ipsilateral sides of fixation, consequently enabling the cancellation of task-unspecific activity affecting both hemispheres (Gratton, 1998).

Notably, a study by Vogel and colleagues (2005) revealed that the CDA can also serve as an indicator of filtering efficiency, evidenced by its amplitude increase when irrelevant items are encoded for storage in VWM, contrary to instructions. To assess filtering efficiency via CDA, researchers employ memory arrays both with and without distractors, facilitating a direct comparison of CDA amplitudes between these conditions (Luria et al., 2016; Vogel et al., 2005). This approach enabled Vogel and colleagues (2005) to demonstrate that individuals with low VWM capacity exhibit similar CDA amplitudes in trials featuring four targets and those with two targets and two distractors. Conversely, high-capacity individuals displayed a CDA amplitude in the filtering condition with two targets and two distractors comparable to a condition with only two target items, indicating superior filtering efficiency.

A more direct measure of VWM's filtering efficiency, such as the filtering cost or CDA, could offer deeper insights into the relationship between VWM and the effect of non-numerical visual cues on numerical acuity. However, despite an extensive literature review, no study has yet explored this potential relationship, indicating a need for further investigation. Additionally, future research should delve into the reasons behind the significant difference in numerical acuity observed between the two rigorously equated trial types generated by the Matlab toolbox CUSTOM algorithm. Specifically, as the present study did not examine the effect of each visual covariate on numerosity discrimination accuracy, further exploration is warranted into the manner in which participants may have relied on non-numerical visual cues during these trials.

The logic underpinning the Matlab toolbox CUSTOM algorithm is that when generating dot arrays with constant total contour or total dot surface area, both configurations are essentially partially (in)congruent, as they incorporate visual parameters that positively and negatively correlate with numerosity. The fact that accuracy was significantly lower on trials with constant total contour in the present study suggests that these trials may have been more incongruent. The reasoning for this stems from the fact that the developers of the Matlab toolbox CUSTOM algorithm demonstrated in their study (De Marco & Cutini, 2020) that trials with equated total contour contain more visual parameters incongruent with numerosity compared to trials with constant total dot surface area.

Specifically, when the total contour is kept constant, three visual parameters – including total surface area, average dot diameter, and classic index of density – correlate negatively with numerosity, whereas only ADBD correlates positively with numerosity (up to numerosities of 30, beyond which it plateaus). In contrast, trials with constant total dot surface area have two visual parameters negatively correlating with numerosity (average dot diameter and ADBD), while only total contour correlates positively with numerosity. This suggests that each visual cue may have a specific weight in influencing numerosity processing, potentially making trials with equated total contour more incongruent. This logic finds support from a recent study using the Matlab toolbox CUSTOM algorithm, demonstrating that numerosity estimation is influenced by the weighting process of both numerosity and non-numerical visual features (Abalo-Rodríguez et al., 2022). Nonetheless, further research is necessary to ascertain the effect or weight of each visual parameter on numerical acuity when utilizing the CUSTOM algorithm.

In conclusion, the results of the empirical investigation included in the present thesis enhance our comprehension of the interplay between numerical acuity and VWM capacity. The results suggested that non-numerical visual cues significantly impact numerical acuity despite the meticulous visual controls implemented by the Matlab toolbox CUSTOM algorithm. The fact that non-numerical visual parameters influence performance in numerosity comparison tasks necessitates future studies investigating pure numerosity processing to employ additional methods beyond the dot-comparison task. Moreover, although significant differences in numerical acuity were observed across trials with varied visual controls, no correlation was observed between this disparity and VWM capacity. This implies that numerosity discrimination and VWM may rely on distinct processing stages, albeit focally suggested to partly overlap. However, the study's limitations, specifically the methodology selected to assess VWM capacity, highlight the need for further research to fully elucidate these relationships. Future investigations could explore alternative indices of VWM capacity that take into account the filtering efficiency of working memory, such as filtering cost or CDA, to better understand the relationship between numerical acuity and VWM. Overall, this research provided a deeper understanding of the cognitive mechanisms underlying numerosity comparison task performance and its interaction with VWM.

References

Abalo-Rodríguez, I., De Marco, D., & Cutini, S. (2022). An undeniable interplay: Both numerosity and visual features affect estimation of non-symbolic stimuli. *Cognition, 222*, 104944.

Aben, B., Stapert, S., & Blokland, A. (2012). About the distinction between working memory and short-term memory. *Frontiers in psychology, 3*, 301.

Adams, E. J., Nguyen, A. T., & Cowan, N. (2018). Theories of working memory: Differences in definition, degree of modularity, role of attention, and purpose. *Language, speech, and hearing services in schools, 49*(3), 340-355.

Agrillo, C., Piffer, L., & Bisazza, A. (2010). Large number discrimination by mosquitofish. *PloS one, 5*(12), e15232.

Allen, R. J., Baddeley, A. D., & Hitch, G. J. (2017). Executive and perceptual distraction in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance, 43*(9), 1677.

Allon, A. S., & Luria, R. (2017). Compensation mechanisms that improve distractor filtering are short-lived. *Cognition, 164*, 74-86.

Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological science, 15*(2), 106-111.

Ashkenazi, S., Haber, H., Shemesh, V., & Silverman, S. (2022). Early Subitizing Development: The Role of Visuospatial Working Memory. *European Journal of Education and Pedagogy, 3*(2), 79-85.

Ashkenazi, S., Henik, A., Ifergane, G., & Shelef, I. (2008). Basic numerical processing in left intraparietal sulcus (IPS) acalculia. *Cortex, 44*(4), 439-448.

Awh, E., Vogel, E. K., & Oh, S. H. (2006). Interactions between attention and working memory. *Neuroscience, 139*(1), 201-208.

Baddeley, A. D. (1986). *Working memory.* Oxford, GB: Oxford University Press.

Baddeley, A. (1998). Recent developments in working memory. *Current opinion in neurobiology, 8*(2), 234-238.

Baddeley, A. (2003). Working memory: looking back and looking forward. *Nature reviews neuroscience, 4*(10), 829-839.

Baddeley, A. (2020). Working memory. In *Memory* (pp. 71-111). Routledge.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. A. Bower (Ed.), *Recent advances in learning and motivation* (Vol. 8, pp. 47–90). New York: Academic Press.

Barton, B., & Brewer, A. A. (2013). Visual working memory in human cortex. *Psychology (Irvine, Calif.), 4*(8), 655.

Bengson, J. J., & Luck, S. J. (2016). Effects of strategy on visual working memory capacity. *Psychonomic Bulletin & Review, 23*, 265-270.

Berryhill, M. E., & Olson, I. R. (2008). The right parietal lobe is critical for visual working memory. *Neuropsychologia, 46*(7), 1767-1774.

Brockmole, J. R., & Logie, R. H. (2013). Age-related change in visual working memory: A study of 55,753 participants aged 8–75. *Frontiers in psychology, 4*, 34741.

Bugden, S., & Ansari, D. (2016). Probing the nature of deficits in the 'Approximate Number System'in children with persistent Developmental Dyscalculia. *Developmental science, 19*(5), 817-833.

Butterworth, B., Varma, S., & Laurillard, D. (2011). Dyscalculia: from brain to education. *Science, 332*(6033), 1049-1053.

Cai, Y., Hofstetter, S., van Dijk, J., Zuiderbaan, W., van der Zwaag, W., Harvey, B. M., & Dumoulin, S. O. (2021). Topographic numerosity maps cover subitizing and estimation ranges. *Nature communications, 12*(1), 3374.

Cappelletti, M., Didino, D., Stoianov, I., & Zorzi, M. (2014). Number skills are maintained in healthy ageing. *Cognitive Psychology, 69*, 25-45.

Castaldi, E., Piazza, M., Dehaene, S., Vignaud, A., & Eger, E. (2019). Attentional amplification of neural codes for number independent of other quantities along the dorsal visual stream. *Elife, 8*, e45160.

Castaldi, E., Piazza, M., & Eger, E. (2021). Resources Underlying Visuo-Spatial Working Memory Enable Veridical Large Numerosity Perception. *Frontiers in Human Neuroscience, 15*, 751098.

Chai, W. J., Abd Hamid, A. I., & Abdullah, J. M. (2018). Working memory from the psychological and neurosciences perspectives: a review. *Frontiers in psychology, 9*, 327922.

Chen, J., Paul, J. M., & Reeve, R. (2022). Manipulation of attention affects subitizing performance: A systematic review and meta-analysis. *Neuroscience & Biobehavioral Reviews, 139*, 104753.

Chen, Q., & Li, J. (2014). Association between individual differences in non-symbolic number acuity and math performance: A meta-analysis. *Acta psychologica, 148*, 163-172.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York: Psychology Press.

Coolen, I. E., Riggs, K. J., Bugler, M., & Castronovo, J. (2022). The approximate number system and mathematics achievement: it's complicated. A thorough investigation of different ANS measures and executive functions in mathematics achievement in children. *Journal of Cognitive Psychology, 34*(6), 796-818.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and brain sciences, 24*(1), 87-114.

Cowan, N. (2008). What are the differences between long-term, short-term, and working memory?. *Progress in brain research, 169*, 323-338.

Cowan, N., Fristoe, N. M., Elliott, E. M., Brunner, R. P., & Saults, J. S. (2006). Scope of attention, control of attention, and intelligence in children and adults. *Memory & cognition, 34*, 1754-1768.

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

Curtis, C. E., & D'Esposito, M. (2003). Persistent activity in the prefrontal cortex during working memory. *Trends in cognitive sciences, 7*(9), 415-423.

Darki, F., & Klingberg, T. (2015). The role of fronto-parietal and fronto-striatal networks in the development of working memory: a longitudinal study. *Cerebral cortex, 25*(6), 1587-1595.

De Beni, R., Pazzaglia, F., Gyselinck, V., & Meneghetti, C. (2005). Visuospatial working memory and mental representation of spatial descriptions. *European Journal of Cognitive Psychology, 17*(1), 77-95.

De Marco, D., & Cutini, S. (2020). Introducing CUSTOM: A customized, ultraprecise, standardization-oriented, multipurpose algorithm for generating nonsymbolic number stimuli. *Behavior Research Methods, 52*, 1528-1537.

De Smedt, B., Noël, M. P., Gilmore, C., & Ansari, D. (2013). How do symbolic and non-symbolic numerical magnitude processing skills relate to individual differences in children's mathematical skills? A review of evidence from brain and behavior. *Trends in Neuroscience and Education, 2*(2), 48-55.

Decarli, G., Paris, E., Tencati, C., Nardelli, C., Vescovi, M., Surian, L., & Piazza, M. (2020). Impaired large numerosity estimation and intact subitizing in developmental dyscalculia. *PLoS One, 15*(12), e0244578.

Dehaene, S. (2001a). Subtracting pigeons: logarithmic or linear?. *Psychological science, 12*(3), 244-246

Dehaene, S. (2001b). Précis of the number sense. *Mind & language, 16*(1), 16-36.

Dehaene, S. (2009). Origins of mathematical intuitions: The case of arithmetic. *Annals of the New York Academy of Sciences, 1156*(1), 232-259.

Dehaene, S. (2011). *The number sense: How the mind creates mathematics.* New York, NY: Oxford University Press.

Dehaene, S., & Cohen, L. (1995). Towards an anatomical and functional model of number processing. *Mathematical cognition, 1*(1), 83-120.

Dehaene, S., Izard, V., & Piazza, M. (2005). *Control over non-numerical parameters in numerosity experiments.* Unpublished manuscript. Retrieved from http://www.unicog.org

DeWind, N. K., Park, J., Woldorff, M. G., & Brannon, E. M. (2019). Numerical encoding in early visual cortex. *Cortex, 114*, 76-89.

Dietrich, J. F., Huber, S., & Nuerk, H. C. (2015). Methodological aspects to be considered when measuring the approximate number system (ANS)–a research review. *Frontiers in Psychology, 6*, 295.

Ditz, H. M., & Nieder, A. (2016). Numerosity representations in crows obey the Weber–Fechner law. *Proceedings of the Royal Society B: Biological Sciences, 283*(1827), 20160083.

Dowker, A. (2023). The componential nature of arithmetical cognition: some important questions. *Frontiers in Psychology, 14*, 1188271.

Drew, T., & Vogel, E. K. (2008). Neural measures of individual differences in selecting and tracking multiple moving objects. *Journal of Neuroscience, 28*(16), 4183-4191.

Ester, E. F., Sprague, T. C., & Serences, J. T. (2015). Parietal and frontal cortex encode stimulus-specific mnemonic representations during visual working memory. *Neuron, 87*(4), 893-905.

Fazio, L. K., Bailey, D. H., Thompson, C. A., & Siegler, R. S. (2014). Relations of different types of numerical magnitude representations to each other and to mathematics achievement. *Journal of experimental child psychology, 123*, 53-72.

Fechner, GT (1860). *Elements of psychophysics* (Vol. 2). Breitkopf and Härtel.

Fecteau, J. H., & Munoz, D. P. (2006). Salience, relevance, and firing: a priority map for target selection. *Trends in cognitive sciences, 10*(8), 382-390.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in cognitive sciences, 8*(7), 307-314.

Feldmann-Wüstefeld, T. (2021). Neural measures of working memory in a bilateral change detection task. *Psychophysiology, 58*(1), e13683.

Flombaum, J. I., Junge, J. A., & Hauser, M. D. (2005). Rhesus monkeys (Macaca mulatta) spontaneously compute addition operations over large numbers. *Cognition, 97*(3), 315-325.

Formoso, J., Barreyro, J. P., Jacubovich, S., & Injoque-Ricle, I. (2017). Possible associations between subitizing, estimation and Visuospatial Working Memory (VSWM) in children. *The Spanish Journal of Psychology, 20*, E27.

Friso-Van den Bos, I., Van der Ven, S. H., Kroesbergen, E. H., & Van Luit, J. E. (2013). Working memory and mathematics in primary school children: A meta-analysis. *Educational research review, 10*, 29-44.

Fuhs, M. W., & McNeil, N. M. (2013). ANS acuity and mathematics ability in preschoolers from low-income homes: Contributions of inhibitory control. *Developmental science, 16*(1), 136-148.

Fukuda, K., Vogel, E., Mayr, U., & Awh, E. (2010). Quantity, not quality: The relationship between fluid intelligence and working memory capacity. *Psychonomic bulletin & review, 17*, 673-679.

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 perfomance XXV, 105.*

Gaspar, J. M., Christie, G. J., Prime, D. J., Jolicœur, P., & McDonald, J. J. (2016). Inability to suppress salient distractors predicts low visual working memory capacity. *Proceedings of the National Academy of Sciences, 113*(13), 3693-3698.

Gebuis, T., Kadosh, R. C., & Gevers, W. (2016). Sensory-integration system rather than approximate number system underlies numerosity processing: A critical review. *Acta psychologica, 171*, 17-35.

Gebuis, T., & Reynvoet, B. (2011). Generating nonsymbolic number stimuli. *Behavior research methods, 43*, 981-986.

Gebuis, T., & Reynvoet, B. (2012a). The interplay between nonsymbolic number and its continuous visual properties. *Journal of Experimental Psychology: General, 141*(4), 642.

Gebuis, T., & Reynvoet, B. (2012b). The role of visual information in numerosity estimation. *PloS one, 7*(5), e37426.

Gebuis, T., & van der Smagt, M. J. (2011). False approximations of the approximate number system?. *PloS one, 6*(10), e25405.

Gilmore, C., Attridge, N., Clayton, S., Cragg, L., Johnson, S., Marlow, N., ... & Inglis, M. (2013). Individual differences in inhibitory control, not non-verbal number acuity, correlate with mathematics achievement. *PloS one, 8*(6), e67374.

Gilmore, C., Cragg, L., Hogan, G., & Inglis, M. (2016). Congruency effects in dot comparison tasks: Convex hull is more important than dot area. *Journal of Cognitive Psychology, 28*(8), 923-931.

Gilchrist, A. L., & Cowan, N. (2014). A two-stage search of visual working memory: Investigating speed in the change-detection paradigm. *Attention, Perception, & Psychophysics, 76*, 2031-2050.

Gratton, G. (1998). The contralateral organization of visual memory: a theoretical concept and a research tool. *Psychophysiology, 35*(6), 638-647.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics* (Vol. 1, pp. 1969-2012). New York: Wiley.

Guillaume, M., Nys, J., Mussolin, C., & Content, A. (2013). Differences in the acuity of the Approximate Number System in adults: The effect of mathematical ability. *Acta psychologica, 144*(3), 506-512.

Guillaume, M., & Van Rinsveld, A. (2018). Comparing numerical comparison tasks: a meta-analysis of the variability of the Weber Fraction relative to the generation algorithm. *Frontiers in Psychology, 9*, 1694.

Hadar, B., Luria, R., & Liberman, N. (2019). Concrete mindset impairs filtering in visual working memory. *Psychonomic bulletin & review, 26*, 1917-1924.

Halberda, J., & Feigenson, L. (2008). Developmental change in the acuity of the" Number Sense": The Approximate Number System in 3-, 4-, 5-, and 6-year-olds and adults. *Developmental psychology, 44*(5), 1457.

Halberda, J., Ly, R., Wilmer, J. B., Naiman, D. Q., & Germine, L. (2012). Number sense across the lifespan as revealed by a massive Internet-based sample. *Proceedings of the National Academy of Sciences, 109*(28), 11116-11120.

Halberda, J., Mazzocco, M. M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature, 455*(7213), 665-668.

Halberda, J., & Odic, D. (2015). The precision and internal confidence of our approximate number thoughts. In *Mathematical cognition and learning* (Vol. 1, pp. 305-333). Elsevier.

Harvey, B. M., & Dumoulin, S. O. (2017). A network of topographic numerosity maps in human association cortex. *Nature Human Behaviour, 1*(2), 0036.

Harvey, B. M., Klein, B. P., Petridou, N., & Dumoulin, S. O. (2013). Topographic representation of numerosity in the human parietal cortex. *Science, 341*(6150), 1123-1126.

Hyde, D. C. (2011). Two systems of non-symbolic numerical cognition. *Frontiers in human neuroscience, 5,* 150.

Inglis, M., & Gilmore, C. (2014). Indexing the approximate number system. *Acta psychologica, 145*, 147-155.

Izard, V., Sann, C., Spelke, E. S., & Streri, A. (2009). Newborn infants perceive abstract numbers. *Proceedings of the National Academy of Sciences, 106*(25), 10382-10385.

Jevons, W. S. (1871). The power of numerical discrimination. *Nature, 3*, 363–372.

Johnson, M. K., McMahon, R. P., Robinson, B. M., Harvey, A. N., Hahn, B., Leonard, C. J., ... & Gold, J. M. (2013). The relationship between working memory capacity and broad measures of cognitive ability in healthy adults and people with schizophrenia. *Neuropsychology, 27*(2), 220.

Jost, K., Bryck, R. L., Vogel, E. K., & Mayr, U. (2011). Are old adults just like low working memory young adults? Filtering efficiency and age differences in visual working memory. *Cerebral cortex, 21*(5), 1147-1154.

Klingberg, T., Forssberg, H., & Westerberg, H. (2002). Increased brain activity in frontal and parietal cortex underlies the development of visuospatial working memory capacity during childhood. *Journal of cognitive neuroscience, 14(*1), 1-10.

Knops, A., Piazza, M., Sengupta, R., Eger, E., & Melcher, D. (2014). A shared, flexible neural map architecture reflects capacity limits in both visual short-term memory and enumeration. *Journal of Neuroscience, 34*(30), 9857-9866.

Kutter, E. F., Bostroem, J., Elger, C. E., Mormann, F., & Nieder, A. (2018). Single neurons in the human brain encode numbers. *Neuron, 100*(3), 753-761.

Kutter, E. F., Dehnen, G., Borger, V., Surges, R., Mormann, F., & Nieder, A. (2023). Distinct neuronal representation of small and large numbers in the human medial temporal lobe. *Nature human behaviour, 7*(11), 1998-2007.

Lee, E. Y., Cowan, N., Vogel, E. K., Rolan, T., Valle-Inclan, F., & Hackley, S. A. (2010). Visual working memory deficits in patients with Parkinson's disease are due to both reduced storage capacity and impaired ability to filter out irrelevant information. *Brain, 133*(9), 2677-2689.

Lee, K., & Cho, S. (2019). Visuo-spatial (but not verbal) executive working memory capacity modulates susceptibility to non-numerical visual magnitudes during numerosity comparison. *Plos one, 14*(3), e0214270.

Leibovich, T., Katzin, N., Harel, M., & Henik, A. (2017). From "sense of number" to "sense of magnitude": The role of continuous magnitudes in numerical cognition. *Behavioral and Brain Sciences, 40*, e164.

Lemer, C., Dehaene, S., Spelke, E., & Cohen, L. (2003). Approximate quantities and exact number words: Dissociable systems. *Neuropsychologia, 41*(14), 1942-1958.

Li, C. H., He, X., Wang, Y. J., Hu, Z., & Guo, C. Y. (2017). Visual working memory capacity can be increased by training on distractor filtering efficiency. *Frontiers in psychology, 8*, 231016.

Libertus, M. E., & Brannon, E. M. (2009). Behavioral and neural basis of number sense in infancy. *Current Directions in Psychological Science, 18*(6), 346-351.

Libertus, M. E., & Brannon, E. M. (2010). Stable individual differences in number discrimination in infancy. *Developmental science, 13*(6), 900-906.

Linden, D. E., Bittner, R. A., Muckli, L., Waltz, J. A., Kriegeskorte, N., Goebel, R., ... & Munk, M. H. (2003). Cortical capacity constraints for visual working memory: dissociation of fMRI load effects in a fronto-parietal network. *Neuroimage, 20*(3), 1518-1530.

Lindskog, M., Winman, A., Juslin, P., & Poom, L. (2013). Measuring acuity of the approximate number system reliably and validly: the evaluation of an adaptive test procedure. *Frontiers in Psychology, 4*, 54144.

Lipton, J. S., & Spelke, E. S. (2003). Origins of number sense: Large-number discrimination in human infants. *Psychological science, 14*(5), 396-401.

Luce, R. D. (1991). *Response times: Their role in inferring elementary mental organization.* Oxford University Press.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390*(6657), 279-281.

Luck, S. J., & Vogel, E. K. (2013). Visual working memory capacity: from psychophysics and neurobiology to individual differences. *Trends in cognitive sciences, 17*(8), 391-400.

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

Luria, R., Balaban, H., Awh, E., & Vogel, E. K. (2016). The contralateral delay activity as a neural measure of visual working memory. *Neuroscience & Biobehavioral Reviews, 62*, 100-108.

Mazzocco, M. M., Feigenson, L., & Halberda, J. (2011). Impaired acuity of the approximate number system underlies mathematical learning disability (dyscalculia). *Child development, 82*(4), 1224-1237.

McNab, F., & Klingberg, T. (2008). Prefrontal cortex and basal ganglia control access to working memory. *Nature neuroscience, 11*(1), 103-107.

Melcher, D., Huber-Huber, C., & Wutz, A. (2021). Enumerating the forest before the trees: The time courses of estimation-based and individuation-based numerical processing. *Attention, Perception, & Psychophysics, 83*, 1215-1229.

Melcher, D., & Piazza, M. (2011). The role of attentional priority and saliency in determining capacity limits in enumeration and visual working memory. *PloS one, 6*(12), e29296.

Menon, V., & Chang, H. (2021). Emerging neurodevelopmental perspectives on mathematical learning. *Developmental Review, 60*, 100964.

Miller, J. (2023). Outlier exclusion procedures for reaction time analysis: The cures are generally worse than the disease. *Journal of Experimental Psychology: General*.

Morey, R. D. (2011). A Bayesian hierarchical model for the measurement of working memory capacity. *Journal of Mathematical Psychology, 55*(1), 8-24.

Mussolin, C., Mejias, S., & Noël, M. P. (2010). Symbolic and nonsymbolic number comparison in children with and without dyscalculia. *Cognition, 115*(1), 10-25.

Nieder, A. (2016). The neuronal code for number. *Nature Reviews Neuroscience, 17*(6), 366-382.

Nieder, A., Freedman, D. J., & Miller, E. K. (2002). Representation of the quantity of visual items in the primate prefrontal cortex. *Science, 297*(5587), 1708-1711.

Nieder, A., & Merten, K. (2007). A labeled-line code for small and large numerosities in the monkey prefrontal cortex. *Journal of Neuroscience, 27*(22), 5986-5993.

Nieder, A., & Miller, E. K. (2003). Coding of cognitive magnitude: Compressed scaling of numerical information in the primate prefrontal cortex. *Neuron, 37*(1), 149-157.

Nieder, A., & Miller, E. K. (2004). A parieto-frontal network for visual numerical information in the monkey. *Proceedings of the National Academy of Sciences, 101*(19), 7457-7462.

Odic, D., & Starr, A. (2018). An introduction to the approximate number system. *Child Development Perspectives, 12*(4), 223-229.

Owen, A. M., McMillan, K. M., Laird, A. R., & Bullmore, E. (2005). N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. *Human brain mapping, 25*(1), 46-59.

Pagano, S., Lombardi, L., & Mazza, V. (2014). Brain dynamics of attention and working memory engagement in subitizing. *Brain Research, 1543*, 244-252.

Park, J., DeWind, N. K., Woldorff, M. G., & Brannon, E. M. (2016). Rapid and direct encoding of numerosity in the visual stream. *Cerebral cortex, 26*(2), 748-763.

Pashler, H. (1988). Familiarity and visual change detection. *Perception & psychophysics, 44*, 369-378.

Passolunghi, M. C., & Lanfranchi, S. (2012). Domain-specific and domain-general precursors of mathematical achievement: A longitudinal study from kindergarten to first grade. *British Journal of Educational Psychology, 82*(1), 42-63.

Peng, P., Namkung, J., Barnes, M., & Sun, C. (2016). A meta-analysis of mathematics and working memory: Moderating effects of working memory domain, type of mathematics skill, and sample characteristics. *Journal of Educational Psychology, 108*(4), 455.

Phillips, W. A. (1974). On the distinction between sensory storage and short-term visual memory. *Perception & Psychophysics, 16*, 283-290.

Piazza, M. (2010). Neurocognitive start-up tools for symbolic number representations. *Trends in cognitive sciences, 14*(12), 542-551.

Piazza, M., De Feo, V., Panzeri, S., & Dehaene, S. (2018). Learning to focus on number. *Cognition, 181*, 35-45.

Piazza, M., Facoetti, A., Trussardi, A. N., Berteletti, I., Conte, S., Lucangeli, D., Dehaene, S., & Zorzi, M. (2010). Developmental trajectory of number acuity reveals a severe impairment in developmental dyscalculia. *Cognition, 116*(1), 33–41.

Piazza, M., Fumarola, A., Chinello, A., & Melcher, D. (2011). Subitizing reflects visuo-spatial object individuation capacity. *Cognition, 121*(1), 147-153.

Piazza, M., Izard, V., Pinel, P., Le Bihan, D., & Dehaene, S. (2004). Tuning curves for approximate numerosity in the human intraparietal sulcus. *Neuron, 44*(3), 547-555.

Piazza, M., Pinel, P., Le Bihan, D., & Dehaene, S. (2007). A magnitude code common to numerosities and number symbols in human intraparietal cortex. *Neuron, 53*(2), 293-305.

Pica, P., Lemer, C., Izard, V., & Dehaene, S. (2004). Exact and approximate arithmetic in an Amazonian indigene group. *Science, 306*(5695), 499-503.

Plebanek, D. J., & Sloutsky, V. M. (2019). Selective attention, filtering, and the development of working memory. *Developmental Science, 22*(1), e12727.

Potrich, D., Sovrano, V. A., Stancher, G., & Vallortigara, G. (2015). Quantity discrimination by zebrafish (Danio rerio). *Journal of Comparative Psychology, 129*(4), 388.

Price, G. R., Palmer, D., Battista, C., & Ansari, D. (2012). Nonsymbolic numerical magnitude comparison: Reliability and validity of different task variants and outcome measures, and their relationship to arithmetic achievement in adults. *Acta psychologica, 140*(1), 50-57.

Revkin, S. K., Piazza, M., Izard, V., Cohen, L., & Dehaene, S. (2008). Does subitizing reflect numerical estimation?. *Psychological science, 19*(6), 607-614.

Riggs, K. J., McTaggart, J., Simpson, A., & Freeman, R. P. (2006). Changes in the capacity of visual working memory in 5-to 10-year-olds. *Journal of experimental child psychology, 95*(1), 18-26.

Robison, M. K., Miller, A. L., & Unsworth, N. (2018). Individual differences in working memory capacity and filtering. *Journal of Experimental Psychology: Human Perception and Performance, 44*(7), 1038.

Rodríguez, C., & Ferreira, R. A. (2023). To what extent is dot comparison an appropriate measure of approximate number system?. *Frontiers in Psychology, 13*, 1065600.

Rottschy, C., Langner, R., Dogan, I., Reetz, K., Laird, A. R., Schulz, J. B., ... & Eickhoff, S. B. (2012). Modelling neural correlates of working memory: a coordinate-based meta-analysis. *Neuroimage, 60*(1), 830-846.

Rotzer, S., Kucian, K., Martin, E., von Aster, M., Klaver, P., & Loenneker, T. (2008). Optimized voxel-based morphometry in children with developmental dyscalculia. *Neuroimage, 39*(1), 417-422.

Rouder, J. N., Morey, R. D., Morey, C. C., & Cowan, N. (2011). How to measure working memory capacity in the change detection paradigm. *Psychonomic bulletin & review, 18*, 324-330.

Schmicker, M., Schwefel, M., Vellage, A. K., & Müller, N. G. (2016). Training of attentional filtering, but not of memory storage, enhances working memory efficiency by strengthening the neuronal gatekeeper network. *Journal of Cognitive Neuroscience, 28*(4), 636-642.

Schneider, M., Beeres, K., Coban, L., Merz, S., Susan Schmidt, S., Stricker, J., & De Smedt, B. (2017). Associations of non-symbolic and symbolic numerical magnitude processing with mathematical competence: A meta-analysis. *Developmental science, 20*(3), e12372.

Sekuler, R., & Mierkiewicz, D. (1977). Children's judgments of numerical inequality. *Child Development*, 630-633.

Smets, K., Sasanguie, D., Szücs, D., & Reynvoet, B. (2015). The effect of different methods to construct non-symbolic stimuli in numerosity estimation and comparison. *Journal of Cognitive Psychology, 27*(3), 310-325.

Sokolowski, H. M., Fias, W., Mousa, A., & Ansari, D. (2017). Common and distinct brain regions in both parietal and frontal cortex support symbolic and nonsymbolic number processing in humans: A functional neuroimaging meta-analysis. *Neuroimage, 146*, 376-394.

Starr, A., DeWind, N. K., & Brannon, E. M. (2017). The contributions of numerical acuity and non-numerical stimulus features to the development of the number sense and symbolic math achievement. *Cognition, 168*, 222-233.

Szűcs, D., Nobes, A., Devine, A., Gabriel, F. C., & Gebuis, T. (2013). Visual stimulus parameters seriously compromise the measurement of approximate number system acuity and comparative effects between adults and children. *Frontiers in psychology, 4*, 444.

Treisman, A., & Zhang, W. (2006). Location and binding in visual working memory. *Memory & cognition, 34*(8), 1704-1719.

Todd, J. J., & Marois, R. (2004). Capacity limit of visual short-term memory in human posterior parietal cortex. *Nature, 428*(6984), 751-754.

Tokita, M., & Ishiguchi, A. (2013). Effects of perceptual variables on numerosity comparison in 5–6-year-olds and adults. *Frontiers in Psychology, 4*, 57899.

Tsouli, A., Harvey, B. M., Hofstetter, S., Cai, Y., van der Smagt, M. J., Te Pas, S. F., & Dumoulin, S. O. (2022). The role of neural tuning in quantity perception. *Trends in Cognitive Sciences, 26*(1), 11-24.

Udale, R., Farrell, S., & Kent, C. (2018). Task demands determine comparison strategy in whole probe change detection. *Journal of Experimental Psychology: Human Perception and Performance, 44*(5), 778.

Unsworth, N., Fukuda, K., Awh, E., & Vogel, E. K. (2015). Working memory delay activity predicts individual differences in cognitive abilities. *Journal of Cognitive Neuroscience, 27*(5), 853-865.

Van Leijenhorst, L., Crone, E. A., & Van der Molen, M. W. (2007). Developmental trends for object and spatial working memory: A psychophysiological analysis. *Child Development, 78*(3), 987-1000.

Van Rinsveld, A., Guillaume, M., Kohler, P. J., Schiltz, C., Gevers, W., & Content, A. (2020). The neural signature of numerosity by separating numerical and continuous magnitude extraction in visual cortex with frequency-tagged EEG. *Proceedings of the National Academy of Sciences, 117*(11), 5726-5732.

Vogel, E. K., & Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature, 428*(6984), 748-751.

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.

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory. *Journal of experimental psychology: human perception and performance, 27*(1), 92.

Wang, L., & Carr, M. (2014). Working memory and strategy use contribute to gender differences in spatial ability. *Educational Psychologist, 49*(4), 261-282.

Whelan, R. (2008). Effective analysis of reaction time data. *The psychological record, 58*, 475-482.

Wilkey, E. D., & Ansari, D. (2020). Challenging the neurobiological link between number sense and symbolic numerical abilities. *Annals of the New York Academy of Sciences, 1464*(1), 76-98.

Wilkey, E. D., Pollack, C., & Price, G. R. (2020). Dyscalculia and typical math achievement are associated with individual differences in number-specific executive function. *Child development, 91*(2), 596-619.

Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition, 74*(1), B1-B11.

Xu, Y., & Chun, M. M. (2009). Selecting and perceiving multiple visual objects. *Trends in cognitive sciences, 13*(4), 167-174.

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature, 453*(7192), 233-235.