UNIVERSITÀ
DEGLI STUDI
FIRENZE

DOTTORATO DI RICERCA TOSCANO IN NEUROSCIENZE

CICLO XXXIII

COORDINATORE Prof.ssa Felicita Pedata

Evidence on perceptual nature of the number sense, on groupitizing phenomenon and on the relationship between number sense, math abilities and math anxiety

Settore Scientifico Disciplinare M-PSI/02

Dottorando
Dott. (Maldonado Moscoso Paula A.)

Tutore
Prof. (Arrighi Roberto)
Prof. (Burr David C.)

Coordinatore
Prof. (*Falicita Pedata)*

Anni 2017/2020

Summary

The present dissertation investigated visual perception of numerosity. In the first part I reviewed the prominent literature about the topic. In the second chapter I described the first experiment, in which I measured confidence and reaction times to study the origins of the well-established visual and motor adaptation effects on numerosity perception. The results reinforce the evidence for a shared mechanism that encodes the quantity of both internally and externally generated events, and shows that the adaptation effects result from changes in sensory encoding, rather than perceptual decisions. More generally, in the study was introduced a novel and useful technique for investigating the mechanisms of numerosity adaptation and sensory adaptation in general. The third chapter investigated the effects of grouping cues on sensory precision of numerosity estimation. The results provide strong evidence that “grouping”, which can improve performance by up to 20%, can be induced by color and/or spatial proximity and occurs in temporal sequences as well as spatial arrays. In the fourth chapter I further examined the groupitizing phenomenon, by testing the hypothesis that the advantage provided by clustering stimuli relies on subitizing. This was achieved by manipulating attention, which is known to strongly affect the subitizing system. In the same chapter I discussed an additional explorative analysis on the relationship between calculation skills and estimation precision of grouped and ungrouped arrays. Taken together, the results showed that groupitizing is truly an attention-based process that leverages on the subitizing system. Furthermore, the outcome of the study suggested that measuring numerosity estimation thresholds with grouped stimuli may be a sensitive correlate of math abilities. In the fifth chapter I went on investigating the neural correlates of the groupitizing phenomenon with both a behavioral and a fMRI study. Similarly to the previous study I measured acuity in estimation of grouped and ungrouped stimuli and additionally I also examined whether the two tasks shared or not the same neural substrate. The results showed that the estimation of grouped and ungrouped stimuli activates similar regions in the right lateralized fronto-parietal network, however, only the presentation of grouped stimuli in the numerosity task elicited the additional activation of regions linked with calculations strategies, for instance the angular gyrus. Moreover, a multivariate pattern analysis showed that parietal activation patterns for individual numerosities could be accurately decoded in the parietal regions independently of the spatial arrangement of the stimuli. Finally, I correlated fMRI decoding accuracy of primary visual areas and angular gyrus with Wfs calculated in the grouped estimation task. Results suggested that the numerical representation in angular gyrus, but not in primary visual areas, is strongly linked with numerical performance and behavior. Overall, the results confirmed psychophysical studies highlighting that groupitizing shares the same regions and neural pattern mechanism of the estimation of ungrouped stimuli, but, furthermore, it also activates brain regions typically activated during calculation tasks. The last part of the dissertation is dedicated to investigating the link between numerosity precision, math abilities and a non-cognitive factor affecting mathematical learning: mathematical anxiety. To this aim, university students with low (< 25th percentile) and high (> 75th percentile) score in the Abbreviate Math Anxiety Scale were tested in multiple domains: a) math proficiency assessed using a standardized test (Mathematics Prerequisite for Psychometrics), b) visuo-spatial attention capacity, measured via a Multiple Object Tracking task, and c) the sensory precision for non-numerical quantities. The results confirmed previous studies showing that math abilities and numerosity precision correlate in subjects with high math anxiety. Furthermore, neither precision in size-discrimination nor visuo-spatial attentional capacity were found to correlate with math capacities. However, within the group with high MA the data also revealed a relationship between numerosity precision and math anxiety, with math anxiety playing a key role in mediating the correlation between participants' numerosity precision and their math achievement. Taken together, this last study suggests an interplay between extreme levels of MA and sensory precision in the processing of non-symbolic numerosity, giving further insight into the processes (and the variables affecting these processes) behind the acquisition of formal mathematical abilities.

In conclusion, the present work assessed the ability to perceive non-symbolic quantities in adults while providing new experimental evidence suggesting its perceptual nature and its link with cognitive and affective factors.

Contents

List of Figures

List of Tables

Publications

Part of the thesis have been included in the following peer-reviewed journal publications.
Experiment in Chapter 2 was included in the following publication. Data collection, analysis and writing were conducted by myself with editing for publication by supervisors:

Maldonado Moscoso, P.A., Cicchini, G.M., Arrighi, R., & Burr, D.C. (2020) Adaptation to hand-tapping affects sensory processing of numerosity directly: evidence from reaction times and confidence. Proc. R. Soc. B 287: 20200801, http://dx.doi.org/10.1098/rspb.2020.0801

Experiment in Chapter 3 was included in the following publication. Data collection were conducted by Anobile, Castaldi and myself. Anobile and Castaldi wrote the research paper with comments and editing provided by supervisors Burr and Arrighi and by myself:

Anobile, G., Castaldi, E., **Maldonado Moscoso, P.A.**, Burr, D.C., & Arrighi, R. (2020). "Groupitizing": a strategy for numerosity estimation. Sci Rep 10, 13436. https://doi.org/10.1038/s41598-020-68111-1

Experiment in Chapter 4 was included in the following publication. Data collection, analysis and writing were conducted by myself with editing for publication by supervisors:

Maldonado Moscoso, P.A., Castaldi, E., Burr, D.C., Arrighi, R., & Anobile, G. (2020). Grouping strategies in number estimation extend the subitizing range. Sci Rep 10, 14979. https://doi.org/10.1038/s41598-020-71871-5

Experiment in Chapter 6 was included in the following publication. Data collection, analysis and writing were conducted by myself with editing for publication by supervisors:

Maldonado Moscoso P.A., Anobile G., Primi C., & Arrighi, R. (2020) Math Anxiety Mediates the Link Between Number Sense and Math Achievements in High Math Anxiety Young Adults. Front. Psychol. 11:1095. doi: 10.3389/fpsyg.2020.01095

Acknowledgments

First and foremost, I would like to thank my supervisors Roberto Arrighi and David Burr who provided me with so much valuable guidance and support throughout the duration of my PhD. I am extremely appreciative of the opportunities they gave me and of the huge amount of time and effort they put into my development as a researcher. I couldn’t have asked for better supervisors.

This research wouldn’t have been possible without the help of all the participants who took part in my studies, so a big thank you goes to those who volunteered to compare hundreds of dot arrays for me.

I would like to thank also Giovanni Anobile, Elisa Castaldi and Guido Marco Cicchini who guided me step by step and patiently explained me everything whenever I needed. A special thanks also to all the good lab mates of the Pisa Vision Lab and Florence Vision Lab.

Thanks also to Mark Greenlee at Regensburg University, who provided me the opportunity to join his team and to learn to conduct an fMRI study. I would like to thank also Markus Becker and Anton Beer for their help and their good company during my stay in Regensburg.

Last, but not least, a special thanks to my family and Irene who supported me during these years.

Declaration

I, the author, declare that the work presented in this thesis is my own and has not been submitted for a degree at any other institution.

Chapter 1

Introduction

Chapter 1

1.1 Numerosity perception

Numbers are an integral part of our everyday life: we use them to count the days in a month, pay for food in a supermarket, and also in complex tasks such as sending an astronaut on the moon. The way numbers are represented and processed has been studied extensively in humans at different developmental stages, as well as in a variety of animal species. One of the most prominent theories in the field of numerical cognition, has been shaped with the contributions of several researches during the last decades (Dehaene, 2011; Gallistel & Gelman, 1992). According to this theory, the ability to process numerical quantities (i.e., the total number of items in a set or its numerosity) is a primary, automatic, and innate ability that can be found across species (Gallistel & Gelman, 2000; Huntley-Fenner, 2001; Jordan & Brannon, 2006; Nieder, 2005). Indeed, in the animal kingdom the ability to rapidly estimate the approximate number of fruits on a tree or the amount of predators/preys in a field is provides a fundamental evolutionary advantage. Experimental studies on a variety of animals (primates, rodents, lions, birds and fishes) have shown that they can discriminate between different non-symbolic quantities (Agrillo et al., 2008, 2009; Bogale et al., 2011; Cantlon & Brannon, 2006; McComb et al., 1994; Meck & Church, 1983; Nieder et al., 2002; Watanabe, 1998). Also, humans are born with the capacity to process non-symbolic numerosities (e.g., dot arrays, groups of objects, number of sounds, etc.). Newborn and infants show the ability to discriminate quantities, as well as being able to engage in rudimentary arithmetic (Brannon et al., 2008; Coubart et al., 2014; Hyde & Spelke, 2011; Izard et al., 2008; Lipton & Spelke, 2003; P. Starkey et al., 1990; Whalen et al., 1999a; Wynn, 1992; F. Xu & Spelke, 2000). Furthermore, according to this theory, symbolic representations of numbers, such as Arabic numerals and number words, which children learn throughout development, are thought to acquire their meaning by being mapped onto the preexisting, non-symbolic representations of number (Piazza, 2010).

Such ability to represent roughly a given quantity seems to be made possible by a core "Approximate Number System" (ANS, Dehaene, 2011). Human adults can reliably compare the cardinality of sets (arrays of dots, sequences of flashes or sound pulses, or motor actions) under conditions that prevent or discourage verbal counting (for example, in dual task situations or when severe time limits are in place) (Barth et al., 2003; Cordes et al., 2001; Whalen et al., 1999a). However, in contrast to precise verbal counting, non-verbal discrimination performance is inaccurate, or noisy. Some indigenous human cultures that lack number words or have a restricted concept of verbal counting rely completely on non-verbal cardinality assessment (Blake, 1991; Gordon, 2004; Pica et al., 2004). For example, the people of the Pirahã tribe in South America have not developed a true number word system, and only use words to designate very small set sizes ('about one'), somewhat larger set sizes ('about two') and sets of many items (a one-two-many system of 'counting'). If asked to match the number of items placed before them with an equal number of objects, the Pirahã show only an imprecise capability to enumerate, with decreasing precision as the numbers become higher. Speakers of Mundurukú, another Amazonian language, lack words for numbers beyond five, but are still able to compare and imprecisely add large sets of items that are far beyond their naming range (Pica et al., 2004). Therefore, humans without a linguistic number concept can only estimate a certain number of items by means of a non-verbal quantification system. To sum it up, then, human adults who lack a verbal counting system, pre-verbal infants and non-verbal animals show an evolutionarily ancient quantification system that operates independently of language (Nieder, 2005).

Discrimination between different numerosities shows a similar response pattern to that of discriminating between different magnitudes, such as brightness, pitch of sound, physical size and weight (Cantlon et al., 2009; but see also Leibovich et al., 2017). In these cases, the ability to detect a change (or to understand that two presented stimuli are different from each other) depends on the ratio between the two to-be-compared magnitudes. For example, it is faster and easier do decide that 10 dots are more numerous than 3 dots (ratio 0.3), than deciding that 10 dots are more numerous than 8 (ratio of 0.8). The ratio effect is thought to result from noisy representation of numerosities (**Figure 1.1A**). It is thought that numerosities are represented in a logarithmic analog format (Dehaene, 2003) where numerosities that have a larger ratio share more representational overlap, and thus cause confusion more easily than numerosities that have a smaller numerical ratio and thus share less representational overlap (**Figure 1.1B**). The ratio between the difference in intensity needed to discriminate two stimuli and their objective intensities is also known as “Weber’s law” (Cantlon et al., 2009). The ratio measurement of individual minimal differences of the intensity between stimuli that can still be discriminated, and their intensity is called Weber fraction. This measure represents an individual’s acuity of numerosity representation: individuals with low Weber fraction scores are able to discriminate much closer numerosities than individuals with high Weber fraction scores. Another index used in the literature to measure the subjective precision is the coefficient of variation (CV), which is the normalization of the standard deviation by the physical stimulus, instead by the perceived (Burr et al., 2013; Lappin et al., 2006; Pomè, Anobile, Cicchini, & Burr, 2019; Testolin & McClelland, 2020). Weber’s law is a characteristic of both nonhuman and human performance and explains the noisy representation of numerosities in the ANS (Cantlon & Brannon, 2006). The processing of features found to obey Weber’s law in classic psychophysical experiments (e.g., loudness, brightness, line-length, etc.) is considered very fast and automatic. Because numerosity processing was found to obey the same law, it has been suggested that numerosity processing is as basic, fast and innate as the processing of brightness, weight, pitch of sound, and so forth (Cantlon et al., 2009; Feigenson et al., 2004; Ross, 2003). This has led to advance the idea that numerosity could reflect a primary visual property of a scene.

On the other hand, several authors have claimed that there is no need to hypothesize a specific system for the processing of numbers, while the approximate number of objects in a scene might be derived by other lower-level proprieties of the image, such as texture density (Allïk & Tuulmets, 1991; Dakin et al., 2011; Durgin, 1995, 2008; Morgan et al., 2014; Tibber et al., 2012). However, this possibility is still subject to debate, and other authors have suggested the existence of specialized mechanisms to perceive numerosity (Anobile et al., 2014; Burr et al., 2018; Burr & Ross, 2008; Cicchini et al., 2016; Kramer et al., 2011; Ross & Burr, 2012, 2010).

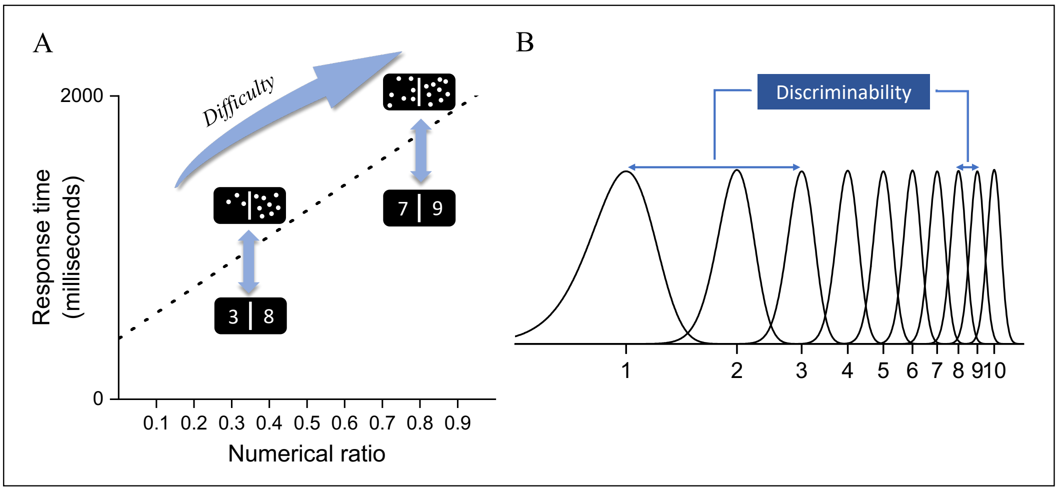


Figure 1.1. Basic effect in numerical cognition. (A) The numerical ratio effect. The plot illustrates the relationship between numerical ratio and response time in a number comparison task. The x-axis describes the numerical ratio: smaller/larger numerosity. The y-axis represents the time it takes to respond to the larger numerosity. Task difficulty increases when the numerical ratio is closer to 1. Inside the plot are examples of symbolic and non-symbolic stimuli: the numerical ratio of 3 and 8 is ~0.37, and the numerical distance is 5; the numerical ratio of 7 and 9 is ~0.77, and the numerical distance is 2. It is also true, then, that task difficulty increases with the decrease in numerical distance. (B) Approximate representation of numerosities. Representation of numbers is thought to be represented on a logarithmic scale. This representation is assumed to be approximate and noisy: larger numbers are represented more approximately, and the representations of adjacent numbers overlap. (Adapted from Leibovich & Ansari, 2016).

1.2 Adaptation and number

One clear sign of the existence of a dedicated perceptual mechanism is its susceptibility to adaptation (Clifford & Rhodes, 2005; Mollon, 1974; Thompson & Burr, 2009). Adaptation is ubiquitous throughout all sensory systems. Adaptation represents a very common kind, throughout the perceptual processing pathways, of experience dependent plasticity, in which our perceptions are "recalibrated" according to the recent history of stimulation, to attune the sensory system to the recent sequence of stimuli and optimize the use of the limited resources of the system (Barlow & Földiák, 1989; Benucci et al., 2013; Boynton, 2004; Kohn, 2007) (although this is not the only role: Ross & Speed, 1991; Solomon & Kohn, 2014).

Number, like most other primary visual attributes, is highly susceptible to adaptation. Brief exposure to either very high or very low numerical quantities changes the apparent numerosity of stimuli subsequently displayed in the same position, causing a repulsive aftereffect whereby the adapted stimulus is perceived as more numerous (overestimation) if the adaptor included fewer dots, and less numerous (underestimation) if the adaptor contained more dots than the adapted stimulus (Arrighi et al., 2014; Burr & Ross, 2008). In his critique of the idea that adaptation acts directly on the abstract representation of numerosity, Durgin (2008) suggested that 'cross-modal studies seem a more promising avenue for distinguishing aftereffects of perceived number from retinotopic aftereffects in the early visual analysis of texture density'. Unlike the visual analysis of texture density, the numerosity adaptation effects are spatially specific. In other words, it is possible to simultaneously adapt different locations of the visual field to high, low or neutral numerosities (Aagten-Murphy & Burr, 2016; Arrighi et al., 2014). Recently, Castaldi et al. (2016) studied the neural effects of adaptation, using functional magnetic resonance imaging (fMRI) techniques. They recorded the BOLD responses to various numerosities from intraparietal sulcus (IPS) and V1 in human observers, before and after they had adapted to an 80-dot stimulus. They used a classifier trained to discriminate between the numerosity of dot clouds before and after adaptation. Importantly, IPS classifiers trained with pre-adaptation presentations could accurately decode number only from other pre-adaptation trials and not from post-adaptation presentations, and vice-versa. This suggests that adaptation changes the cortical maps underlying the presentation of numerosity in IPS, and not in the early stages of analysis, as have been suggested by Durgin (2008). Arrighi et al. (2014) showed that numerosity adaptation occurs not just with the classic cloud of dots, but also with sequentially presented stimuli. Adapting to a slow presentation rate (2 Hz) caused an overestimation of the number of disks subsequently presented. On the other hand, adapting to a fast sequence (8 Hz) produced an underestimation of the subsequent stimuli (**Figure 1.2A**). Arrighi et al. went on to examine cross-modal adaptation to numerical sequences. They adapted to a stream of sounds and asked participants to estimate the number of visual flashes, and they also changed the apparent numerosity of a series of sounds after visual flashes adaptation. Lastly, they showed that adaptation occurs independently of the format used to display numbers. They tested the effect of adaptation to a stream of flashes on “classic”, simultaneously presented cloud of dots. In all of these conditions, they found that the magnitude of the adaptation effect is completely comparable to the previous within-modality experiment, finding an effect of under- and over-estimation (depending on the numerosity of the adaptor) of about the same extent of the previous experiments (**Figure 1.2B**). Like adaptation to spatial numerosity, the temporal numerosity aftereffect was also spatially selective. All these results point to the existence of a very generalized number sense, transcending space, time, and sensory modality.

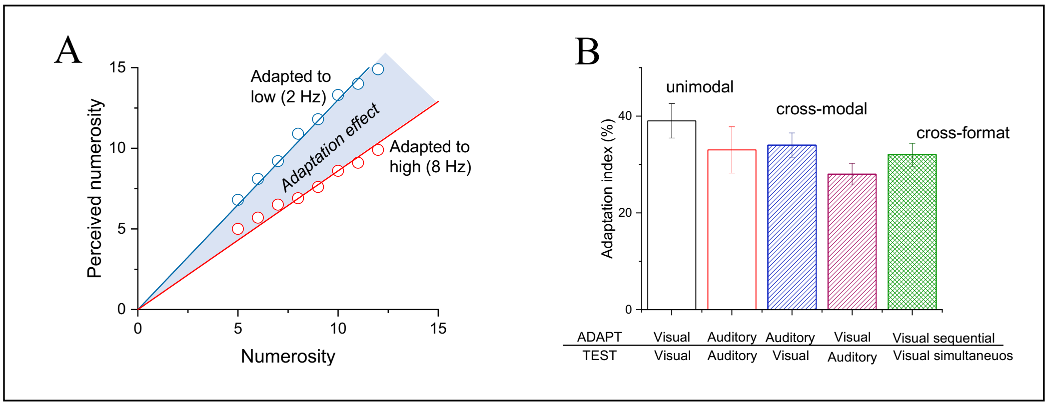


Figure 1.2. Cross-format and cross-modal numerosity adaptation. (A) Sample results in the visual unimodal adaptation task. Adapting to low numbers (2 Hz, blue circles) produces an overestimation of numerosity and adaptation to high numbers (8 Hz, red circles) an underestimation. Data were well-fitted with linear regressions (lines on the data) forced to pass throughout zero. The shaded region represents the strength of the adaptation effect (adaptation index) given by the difference in slope of the regression lines. (B) Mean adaptation indexes for the various experimental conditions were: uni-modal visual and auditory adaptation; cross-modal auditory-visual and visual-auditory adaptation; and “cross-format” adaptation (adapting to serial presentation, testing with simultaneous). Bars show average data, error bar represent ±1 s.e.m. (Reproduced with permission from Arrighi et al., 2014)

Neurophysiological evidences from macaque monkeys have suggested that numerosity can be important for the generation of actions (Sawamura et al., 2002, 2010). To test the link between numerosity and actions, Anobile et al., (2016) investigated whether adapting to actions could affect number perception. Subjects performed rapid or slow finger tapping and then judged the numerosity of sequences of flashes, or of arrays of dots. As with previous results, adapting to slow tapping caused overestimation, and adapting to fast tapping caused underestimation. Again, adaptation works equally well both for sequences of flashes and for clouds of dots (**Figure 1.3A and 1.3B**), and it also affects the apparent numerosity of auditory sequences **(Figure 1.3C and 1.3D**; Togoli et al., 2020). And just as the temporal adaptation is selective in spatiotopic rather than retinotopic coordinates, adaptation to tapping is selective depending on the spatial position of the tapping hand, not on which hand does the tapping. Under the three conditions tested (right hand tapping right and left, and left hand tapping left) the adaptation effects were strong only when the hand (either left or right) was tapped on the same side as the stimuli were presented. This result is important as it reveals the interplay between action and perception in the numerical dimension, potentially underlying successful interaction between our body and objects in the environment, for instance when planning the number of movements to execute based on the number of elements in our peri-personal space.

It has recently been questioned whether adaptation reveals changes in cognitive decisional processes rather than changes in neural mechanisms (Firestone & Scholl, 2016; Morgan et al., 2011). In the first part of the present work, it will be illustrated and discussed an experiment aimed to address the question about the nature of the core mechanism underlying the numerosity adaptation. To disentangle between the hypothesis that numerosity adaptation is a perceptual phenomenon occurring via recalibration of the tuning of numerosity mechanisms against the idea it just reflects a bias in the decisional processes, I studied visual and motor adaptation of numerosity perception while measuring confidence and reaction times.

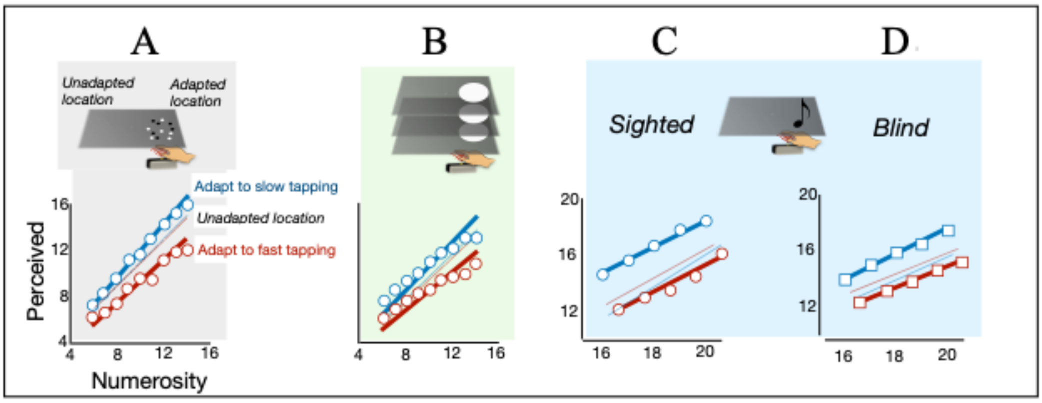


Figure 1.3. Motor adaptation effect on numerosity perception. (A-D) Participants tapped rapidly or slowly in mid-air while keeping their eyes fixed on a central point on a blank monitor (motor adaptation). After the adaptation phase (~6s) a stimulus was briefly (~250 ms) presented around the motorically adapted region, or in the opposite hemifield. After a sequence of fast tapping, the perceived numerosity of dots (A), flashes (B) and sounds (C-D) were all underestimated compared to what happened with slow adaptation (red compared with blue symbols). The motor adaptation effects were all spatially selective, occurring only for stimuli presented around the adapted location (continuous lines) not extending to stimuli presented in the opposite hemifield (dotted lines). The effect on auditory numerosity also occurred in congenitally blind adults (D). (Reproduced with permission from Anobile, Arrighi, et al., 2016; Togoli et al., 2020).

1.3 Mechanisms behind the numerosity perception

Anobile et al. (2016) suggested that there exist three different regimes in number analysis (**Figure 1.4**).

A particular aspect of numerical perception is the so called subitizing. It is the capacity to rapidly and accurately enumerate a small number of items (1-3 or 4). The term subitizing (from the Latin “subitus” which means suddenly) was coined by Kaufman & Lord (1949). It has been historically demonstrated that when participants were asked to enumerate visual sets of items their enumeration time was almost constant up to 4 items (by about 40-100 ms/item). A similar trend was also found when researchers measured the performance precision (Dehaene, 2011; Whalen et al., 1999b). Subitizing is robust, and occurs for both sequential and simultaneous stimuli, in all sensory modalities (Anobile, Arrighi, et al., 2019; Butterworth, 2019; Camos & Tillmann, 2008; Dehaene, 2011; Plaisier et al., 2009). Lastly, subitizing is highly dependent on attention (Anobile et al., 2012; Anobile, Tomaiuolo, et al., 2020; Burr et al., 2010, 2011; Egeth et al., 2008; Olivers & Watson, 2008; Pomè, Anobile, Cicchini, Scabia, et al., 2019; Railo et al., 2008; Vetter et al., 2008; Xu & Liu, 2008).

As I mentioned above for higher numbers, numerosity is estimated slowly (with an increase by about 250-350 ms/item) and through an error-prone process, where error increases linearly with numerosity, following Weber's law (Atkinson et al., 1976; Kaufman et al., 1949; Mandler & Shebo, 1982; Pomè, Anobile, Cicchini, & Burr, 2019; Ross, 2003). Weber fractions vary considerably between individuals, but also with eccentricity, being lower for centrally-viewed rather than peripherally-viewed stimuli (Anobile et al., 2014). Studies which manipulated attentional resources during numerosity tasks found that attentional costs remain constant over the estimation range (Burr et al., 2010; Shapiro et al., 1997). At high densities, texture-like mechanisms kick in. While it seems to be equally sensitive to all eccentricities, the reaction times decrease at a lower numerosity for the more eccentric stimuli (Pomè, Anobile, Cicchini, & Burr, 2019). In this regime, Weber fraction is not constant, but decreases with the square root of density. Importantly, the transition from numerosity to texture is determined by density, defined as average center-to-center spacing of the elements (Anobile et al., 2014).

There are unlikely to be strict demarcations between the regimes. As suggested by Burr and colleagues (2010, 2011) the numerosity range extends well into the subitizing range; but, when measuring thresholds, the most sensitive mechanism (subitizing) prevails, resulting in an errorless performance. However, when subitizing is compromised by diminishing attention, even for a low number of elements (1-4) estimation yields similar Weber fractions to the rest of the estimation range, and also adaptation occurs when attention is diminished. Similarly, when asked to make numerosity or density judgements between stimuli of equal area, one can act as a proxy to the other, and the most sensitive will prevail.

In the third and fourth chapter of the present work, I will illustrate and discuss two experiments that provide a genuine interconnection between subitizing and estimation mechanisms.

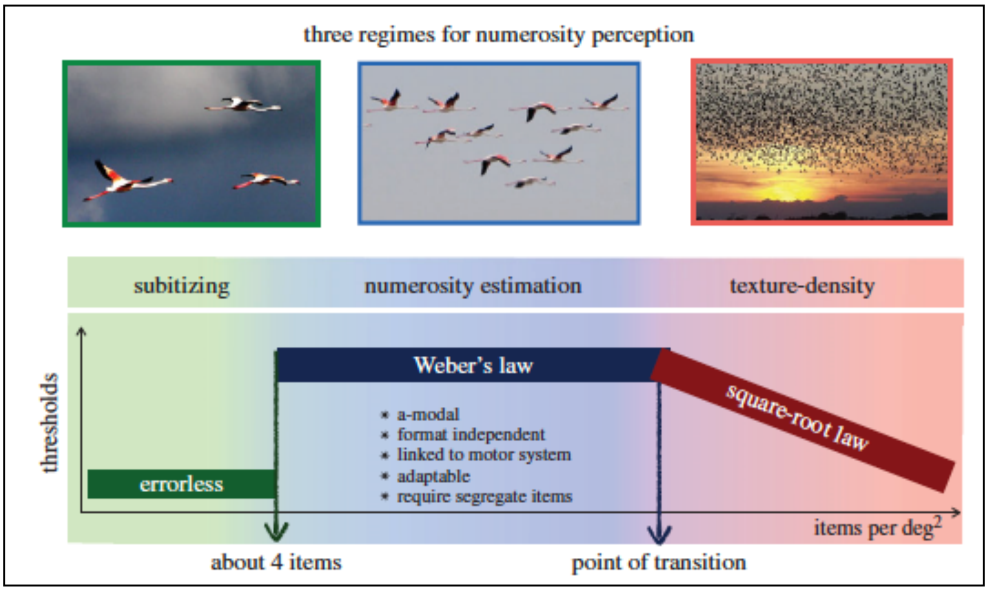


Figure 1.4. **Illustration of the three regimes of numerosity perception**: subitizing, estimation and texture. (Reproduced with permission from Anobile, Cicchini, et al., 2016).

1.4 Neural correlates of the ANS and arithmetic calculation

Most of the current neuroscience research on numbers has focused on the neural correlates of the representation of numerical quantities in the brain. Single-cell recordings in macaque monkeys have identified neurons tuned to specific numerosities of visual arrays in the posterior parietal cortex and prefrontal cortex (Nieder et al., 2002; Nieder & Miller, 2004; Roitman et al., 2007; for review see: Nieder, 2016).

One of the most influential models of how the numerical information can be represented in the human brain is the triple-code model (Dehaene et al., 2003). Based on behavioral, neuroimaging and neuropsychological studies, this model proposed that at least three different types of interconnected representations exist of number. The first is a quantity system, a representation of the size and distance relations between numbers. The second is a visual representation in which numbers can be coded as strings of Arabic digits (visual system) and the last a semantic system where numbers are represented lexically, phonologically, and syntactically. The triple-code model proposes that three cerebral areas may be recruited during number processing: the bilateral intraparietal sulcus (IPS) which encodes the abstract representation of numerical magnitude; the left angular gyrus (AG), associated with verbal processing of numbers and the bilateral posterior superior parietal areas (SPL) associated with spatial and nonspatial attention. Neuropsychological observations show a double dissociation between the functions supported by IPS and left angular gyrus, suggesting that the neural bases of calculation are heterogeneous. For instance, in patients with left parietal lesions (within the angular gyrus) and/or atrophy showed most often impairment in multiplication performance and semantic representation of numerical quantities but not, or to a minor extent, addiction and/or subtraction performance (Dehaene & Cohen, 1997; Delazer & Benke, 1997; Van Harskamp & Cipolotti, 2001; Whalen et al., 1997). Conversely, observations of lesion or cortical stimulation intra-operatively within left IPS, disrupted subtraction but not multiplication (Duffau et al., 2002; Van Harskamp & Cipolotti, 2001). These findings suggest that the left angular gyrus may support multiplications by recalling the solution from verbal memory, while IPS may be support subtractions through some sort of internal manipulation of numerical quantities on an internal number line, probably similar to the strategy employed to solve numerical comparisons (Dehaene et al., 2003).

The early neuroimaging studies were mostly performed using low resolution and whole brain averaging analysis. More recently, various techniques have been used to provide a more detailed and finer scale description of numerical representation. The first attempt to obtain results more similar to those obtained from single neuron recordings in monkey used fMRI habituation. Piazza and collogues (2004) habituated participants to a constant number of items while varying stimuli low-level features (e.g. dot size, cumulative area and spacing, overall luminance and density): a change in numerosity lead to a release from adaptation in bilateral intraparietal cortex, the size of which reflected the ratio between adaptation and deviant number, in agreement with Weber's law (Cohen Kadosh et al., 2011; Demeyere et al., 2014; He et al., 2015; Jacob & Nieder, 2009; Piazza et al., 2004, 2007; Roggeman et al., 2011). Using this technique the authors reported tuning curves similar to those described in the macaque monkeys. Habituation signals have been recorded from the parietal cortex of infants and children (Cantlon & Brannon, 2006; Hyde & Spelke, 2011; Izard et al., 2008), suggesting that numerosity perception is a very primordial processing preceding language and explicit numerical learning.

To further investigate numerical representation at finer spatial scale, multivariate pattern analysis (MVPA), population receptive field (pRF) and high-resolution functional imaging have been used. With the use of MVPA several experiments have demonstrated that it is possible to decode numerosity in the intraparietal sulcus (Bulthé et al., 2014; Castaldi et al., 2016; Damarla & Just, 2013; Eger et al., 2015; Eger et al., 2009). Eger and collogues (2009) were the first to successfully decode symbolic and non-symbolic numerosities from the pattern of activity read out from the parietal regions.

pRF analysis combined with high- resolution functional imaging has also allowed researchers to measure topographic numerosity maps in humans. These maps, where individuals voxels respond preferentially to different numbers of visual items, were found to be located superior/medially in the superior parietal lobule (Harvey et al., 2013). These maps are mostly (although not completely) overlapping with the visual field map representations: a series of retinotopic visuals field maps identified using phase-encoded mapping (Konen & Kastner, 2008; Sereno et al., 2001; Silver et al., 2005; Swisher et al., 2007), labelled from IPS0 (the most posterior) to IPS 5 (the most anterior) (Konen & Kastner, 2008; Silver et al., 2005; Swisher et al., 2007).

Recently, Castaldi et al. (2019) provided direct evidence for a sensory mechanism capable of differentiating signals related to numerosity from those related to associated non-numerical dimensions from early stages of cortical processing on, which can be independently and progressively amplified across the dorsal visual stream when numerical information is explicitly task-relevant.

High-resolution neuroimaging has also allowed researchers to investigate the sub-regional specialization of IPS. Specifically, it has been found that that more medial parts of IPS are preferentially recruited during viewing of non-symbolic (over symbolic) numerical stimuli, while more lateral parts of IPS are preferentially recruited during numerical operations (comparison and calculation) (Castaldi, Vignaud, et al., 2020).

Although the parietal cortex is fundamental to number processing and calculation, other regions are involved as well (Ischebeck et al., 2009; Zago et al., 2008; Zhou et al., 2007). In a metanalysis, Arsalidou and Taylor (2011) showed that prefrontal cortex, particularly in middle and superior frontal gyri, is essential for number and calculation. They also showed that cingulate gyri and insula are active in various numerical tasks and calculation. Overall, these studies suggest that the network for numerosity processing and calculation involves a larger network that extends beyond the parietal cortex.

Chapter five describes an fMRI study in which I investigated the neural correlates of numerosity perception when stimuli are presented in arrays of ungrouped and grouped items.

1.5 Relation between the approximate number system and math abilities

How do we develop mathematical competence? Is there a relationship between the ANS and math abilities? In recent years there has been substantial interest in addressing these questions by investigating individual differences in children and adults' abilities when performing basic arithmetic operations and non-symbolic discrimination tasks. Some researchers have hypothesized that the ANS is a cornerstone for the development of mathematical abilities (Piazza, 2010). Halberda, Mazzocco and Feigenson (2008) found that number acuity remained a significant predictor of performance in standardized mathematical achievement tests, even when the effect of several other cognitive processes (e.g. intelligence, verbal IQ, visual working memory, gender) was statistically controlled. A good deal of other evidence shows that numerosity discrimination thresholds are a reliable predictor of both current and future math achievements in school-age children (Anobile et al., 2013; Anobile, Arrighi, et al., 2018; Bonny & Lourenco, 2013; De Smedt et al., 2009; Inglis et al., 2011; Libertus et al., 2011, 2013; Mazzocco et al., 2011; Starr et al., 2013) and adults (Halberda et al., 2012; Libertus et al., 2012; Lourenco et al., 2012; Lyons & Beilock, 2011). A recent large-scale meta-analysis of multiple longitudinal data sets concluded that math ability at school entry is the strongest predictor of later school achievement (Duncan et al., 2007). Training studies on non-symbolic approximate number tasks show improvements in symbolic arithmetic performance in preschoolers (Park et al., 2016), school-age children (Hyde et al., 2014; Räsänen et al., 2009; Wilson et al., 2006; Wilson et al., 2006, 2009) and in adults (Park & Brannon, 2013, 2014). Accordingly, children with dyscalculia, a neurodevelopmental disorder affecting mathematical and numerical learning, often exhibit a reduced performance in comparing non-symbolic quantities with higher Weber fractions compared to typically developing children (Anobile, Cicchini, et al., 2018; Mazzocco et al., 2011; Piazza et al., 2010). Therefore, an accurate representation of non-symbolic numerical quantities constitutes an important predictor of later mathematical achievements. A "noisy" representation of non-symbolic numerical quantities can compromise the acquisition of subsequent numerical skills and math achievements. A study conducted by Anobile and colleagues (2018) found that math reasoning has a specific relationship with the encoding of spatial information about quantity in children. Here we assessed math abilities in a sample of children and adults, and we correlated it against the ability to discriminate as well to estimate the pattern of spatial and temporal stimuli. The correlation analysis between the ANS tasks (discrimination between cloud of dots, estimation of dots, estimation of sequences of flashes and sequences of sounds) revealed that these measures correlate each other in children, but not in adults. The experiment replicated previous studies showing that children with higher precision in discriminating and estimating simultaneous visual numerosity show higher abilities in formal math (**Figure 1.5A and 1.5B**; Feigenson et al., 2013; Halberda et al., 2008; Libertus et al., 2013; Piazza et al., 2010). Interestingly, the precision in estimating sequential numerosity (sequences of flashes or sounds) was completely unrelated to math abilities in children (**Figure 1.5C and 1.5D**). These results are in line with the idea that human mathematical thought arises from the cultural recycling of ancient brain areas representing features more naturally linked to math concepts, such as visual space (Dehaene, 2011; Dehaene et al., 1999).

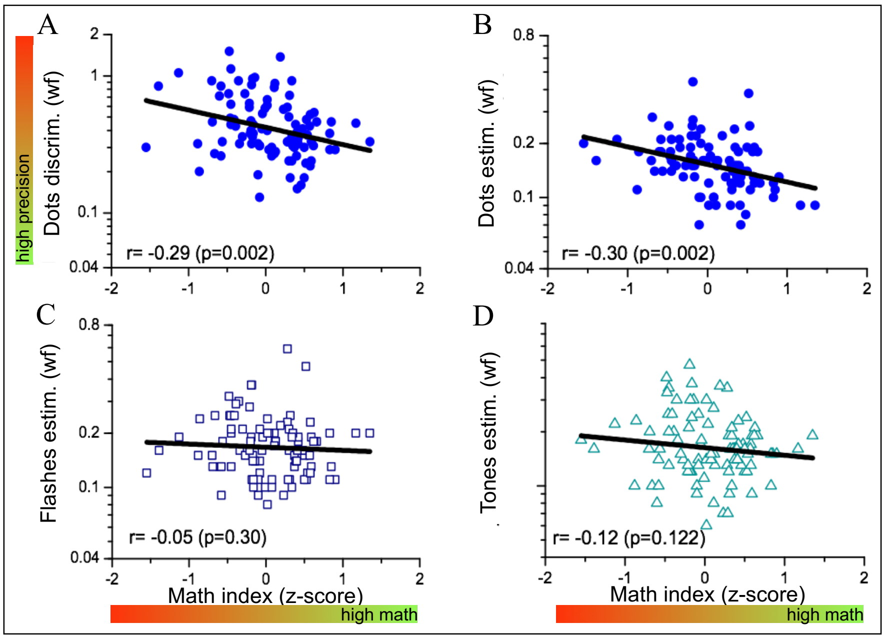


Figure 1.5. Correlations between ANS and math skills in children. ANS Weber fractions are plotted against standardized math skills level for the four ANS tasks: (A) spatial ensemble discrimination, (B) spatial estimation, (C) estimation of flashes sequences, and (D) estimation of tones sequences. Filled symbols report statistically significant correlations (Pearson zero-order correlations with alpha-level = 0.05/15 = 0.0033). (Reproduced with permission from Anobile, Arrighi, et al., 2018).

Although the above mentioned studies strongly support a causal link between ANS and math capacity, the evidence in the literature remains controversial (Lindskog & Winman, 2016). Several studies have not found a correlation between the ANS precision and mathematical abilities in children (Holloway & Ansari, 2009; Lonnemann et al., 2011; Lyons et al., 2014; Rousselle

& Noël, 2007; Sasanguie et al., 2012, 2014) and in adults (Anobile, Arrighi, et al., 2018; Castronovo & Göbel, 2012; Inglis et al., 2011; Price et al., 2012), while others found that training in approximate numerosity does not change formal math abilities (Obersteiner et al., 2013; Sullivan et al., 2016). It is important to point out that the existing body of studies has typically employed standardized or curriculum measures of mathematical achievement, which encompass a range of mathematical skills (e.g. number fact knowledge, conceptual understanding, strategy use and proficiency). Any meaningful relationship between numerical magnitude representations and mathematics will likely vary across different mathematical skills. In other words, numerical magnitude processing will be more important for some aspects of mathematical competencies than others (De Smedt et al., 2013). The inconsistencies in the results found in literature might be also explained by differences in the age of the participants, the stimuli used, indices that were calculated to tap into non-symbolic processing and mathematics achievement as well as by emotional states that could influence subject's performance. In Chapter 4 I will present a pilot study which suggests a possible explanation of the discrepancies in the literature about the relationship between ANS and math abilities. Lastly, in chapter 6 I will show that even participants' emotional states play a role in mediating the link between the ANS and math achievement.

Chapter 2

Adaptation to hand-tapping affects sensory processing of numerosity directly: evidence from reaction times and confidence

Chapter 2

2.1 Introduction

Perceptual adaptation is a form of short-term plasticity, usually generated by observing for some time a particular stimulus, such as a steadily drifting pattern. Adaptation has proven to be a fundamental psychophysical tool to study many perceptual properties, including high-level properties such as face identity and expression (Leopold et al., 2005; Mollon, 1974; Thompson & Burr, 2009). It has also proven invaluable in the study of the perception of *numerosity*, bringing this field of cognitive research into the realm of perceptual research (Burr & Ross, 2008; Butterworth, 2008; Castaldi et al., 2016). Recently, cross-modal and cross-format adaptation have been used to demonstrate a "generalized sense of number", showing strong interactions between the numerosities of spatial arrays of objects and temporal sequences of events (Arrighi et al., 2014). Even more intriguingly, the authors went on to show interactions between numerosity perception and motor action: fast tapping reduces the apparent numerosity of both temporal sequences and spatial arrays, while slow tapping has the opposite effect (Anobile, Arrighi, et al., 2016).

These results are clearly important as they point to specific neural interactions between different forms of numerosity representation, reinforcing the neurophysiological evidence reported in macaque monkeys (Nieder & Dehaene, 2009). They also show strong neural links between numerosity and motor action, again with parallels in the neurophysiological literature (Sawamura et al., 2002). But do adaptation studies truly reveal underlying neural mechanisms as Mollon (1974) claimed ("if you can adapt it it's there")? Can we think of adaptation as the "psychologists microelectrode", as suggested by Frisby (1979)?

It has recently been questioned whether adaptation necessary reveals underlying neural mechanisms, with suggestions that they could result from changes in observer criteria, driven by cognitive, decisional processes, particularly for certain "high-level" aftereffects (for discussion see Firestone & Scholl, 2016). To demonstrate this possibility, (Xia et al., 2016) et al. (2011) showed that observers could simulate the effects of adaptation by adopting simple decision rules, along the lines of "if unsure say fewer". This strategy resulted in a clear shift of psychometric functions, without broadening the width of the functions (reflecting preserved precision). Therefore, it is possible that in the numerosity adaptation experiments the changes in the psychometric functions do not reflect changes in neural representations of number, but in a cognitive, decision strategy in reporting numerosity. Possibly after rapid tapping there is a tendency to report uncertain numerosities as low, and after slow tapping to report these as high. This could conceivably account for the changes in apparent numerosity, without invoking the action on neural mechanisms.

Morgan et al.'s idea can be illustrated with a simple simulation shown in **Figure 2.1**. The red curve illustrates a typical psychometric function, modelled by a cumulative Gaussian error function. The blue curve illustrates a hypothetical function of subjective confidence, based on the consistency of participant responses: one when certain, zero when guessing. On the basis of data from this study (see **Figure 2.3**) we assume minimal confidence is 50%, but this is not essential to the demonstration. Confidence should be minimal at the point of subjective equality, where sensory information is least. The green curve is the simulation of the strategy "if unsure say 'fewer'" (the product of the two probability functions), causing a downward shift of the curve, which necessarily shifts the function rightwards. The downward shift in the curve is virtually indistinguishable from a rightward shift caused by sensory adaptation to numerosity. However, if it is confidence that drives the downward shift, the confidence function itself should not change, but remain centred at the PSE of the unadapted function.

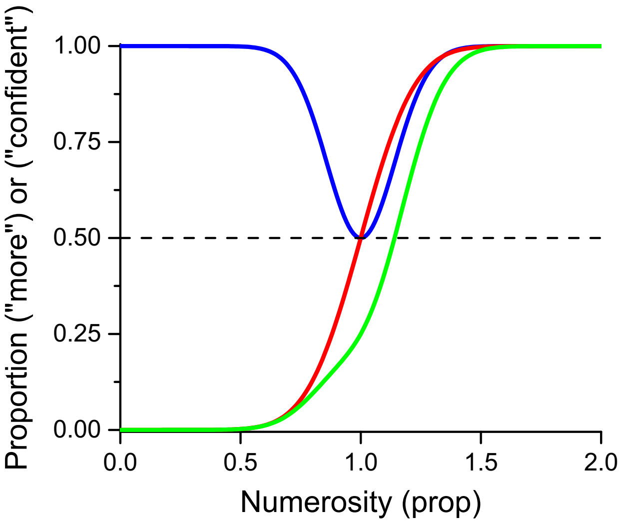


Figure 2.1. Simulation of psychometric function. Simulation showing how response biases could induce a shift in psychometric function resembling a real sensory change. The red curve shows a hypothetical psychometric function for a numerosity discrimination task. The blue curve plots confidence level based on the relative numerosity difference between the stimuli. The green curve shows the result of a decision strategy "less if unconfident", obtained by the pointwise product of two functions.

Gallagher and colleagues (2019) took advantage of this fact to propose a novel way of distinguishing between sensory effects in adaptation and higher-level decisional biases, based on the assumption that confidence in the perceptual decision will scale with the strength of sensory evidence. In the typical two-alternative matching experiment used to measure adaptation, where participants choose which of two stimuli was the largest, the strength of sensory evidence will be weakest when their internal representations of magnitude are the same: that is, at the point of subjective equality (PSE). Therefore, the PSE should also correspond to the point of minimal confidence. If the PSE shifts with adaptation-induced changes in internal representations of magnitude, the shift in PSE should be accompanied by a comparable shift in minimal confidence. If, on the other hand, the adaptation results from weak confidence and a decision rule (as simulated in **Figure 2.1**), the confidence ratings should remain minimal at the point of physical equality, and not shift with adaptation. Gallagher et al. (2019) showed that adaptation to visual motion shifted not only the point of perceived equality of motion, but also the point of maximal decisional uncertainty. On the other hand, instructing participants to introduce a systematic response bias (along the lines of replicating Morgan et al.'s experiment) did not shift the point of maximal uncertainty.

Another common tool in sensory research is reaction-times, which also vary systematically with sensory strength, well approximated by a power function of the stimulus strength plus a constant (Piéron's law: Piéron, 1914). Following the same logic discussed above, reaction-times should also vary on a two alternative forced choice task, being maximal when the sensory representations of the two are most similar, at the point of subjective equality. Therefore, adaptation should also shift the peak in reaction-times, following the shift in PSE, if the effects are sensorial rather than decisional. If they remain anchored at physical equality, the adaptation is more likely to reflect response or decision biases.

In this study we investigate how adaptation to numerosity affects confidence ratings and reaction-times. We study two types of adaptation: visual adaptation to dense dot arrays (Burr & Ross, 2008), and motor adaptation to fast and slow hand-tapping (Anobile, Arrighi, et al., 2016). The results show that both types of adaptation cause concomitant changes in both minimal confidence and maximal reaction-times, suggesting that the effects of both adaptation to high-numerosity and to manual tapping are sensory rather than biases in decision.

2.2 Methods

Stimuli were presented on an Acer LCD monitor (screen resolution of 1920×1080, refresh rate 60 Hz) subtending 50°×29° at the subject view distance of 57 cm. They were created with PsychToolbox routines for MATLAB (ver. R2016a, the Mathworks, Inc.) on a PC computer running Windows 7. In the motor adaptation conditions, hand movements were monitored by an infrared motion sensor device (Leap motion controller – https://www.leapmotion.com) running at 60 Hz.

We used a standard forced-choice paradigm (**Figure 2.2**). Stimuli were brief (250 ms) patches of dots, presented sequentially to the left and right of fixation, with a 200 ms pause between them. Each patch covered a circular region of 8° in diameter, centred at 7° from screen centre. Dots were 0.3° diameter, separated from each other by at least 0.25°, half white and half black (to balance luminance), presented on a grey background. The patch to the left of fixation was the reference, with numerosity fixed at 16 dots; that to the right was the probe, with numerosity varying randomly from 8 to 32 dots (numerosity drawn from linear rectangle distribution). Participants first judged whether the stimulus on the left or the right appeared more numerous, then indicated their confidence in the judgments by pressing the up or down arrow (low or high confidence respectively). We also measured the reaction-times of the numerosity judgments, and report the mean, after removing outliers (more ±3 standard deviations from the mean).

2. 2. 1 Adaptation

For the visual adaptation experiment, 12 participants (11 naïve to the purpose of the study and 1 author; mean age 28 with normal or corrected-to-normal vision) adapted to an array of 60 dots (adapt to high) at the same position as the probe stimulus, for 40 s at the beginning of each session, then for 6 s top-up periods. Stimuli were presented 1 s after adaptation. Each participant performed a total of 432 trials. For the adaptation-to-tapping experiment, participants (9 naïve to the purpose of the study and 1 author; mean age 28 with normal or corrected-to-normal vision) made a series of hand-tapping movements (pivoting at the wrist) on the right side of the screen until a white central fixation point turned red (the stop signal); 1 s later the stimuli were presented. In one condition participants tapped as rapidly as possible, in another at around 1 Hz. The program continuously monitored tapping via the infrared motion sensor: if a tap occurred after the presentation of the test stimulus, the trial would be aborted. After the stimuli presentation, subjects were required to press left arrow when the stimulus at left was perceived as more numerous, or right arrow when the righthand stimulus was perceived as more numerous. They then pressed up-arrow if they were confident about the numerosity response or down-arrow if they were not. Participants were unaware that we also measured the reaction-time of the numerosity response, and they were not explicitly asked to make speeded responses. Three blocks of 24 trials were run for each condition.

2. 2. 2 Manipulation of rewards

We devised a control experiment to compare with adaptation, where we manipulated the reward rules. 10 adults participated in this study, 9 naïve to the purpose of the study (mean age 28 with normal or corrected-to-normal vision). Here there was no adaptation, but participants played a point-based game, with three types of reward regimes (in different blocks). In baseline blocks, they received 1 point for each correct response and lost 1 for every error (performing on average at 85% correct). In "reward-low" blocks, they received 2 points for correctly responded "less than", and lost 1 for each error; and in "reward-high", 2 points for correctly responding "greater than", losing 1 for an error. They also indicated by pressing up-arrow if they were confident about the numerosity indicated was "less" or "greater than" or down arrow if they were not. They were given feedback on earning 50 points, and again at 80 points. Three blocks with at least 79 trials were run for each condition. We also measured the reaction-time of the response, and again participants were not explicitly asked to make speeded responses.

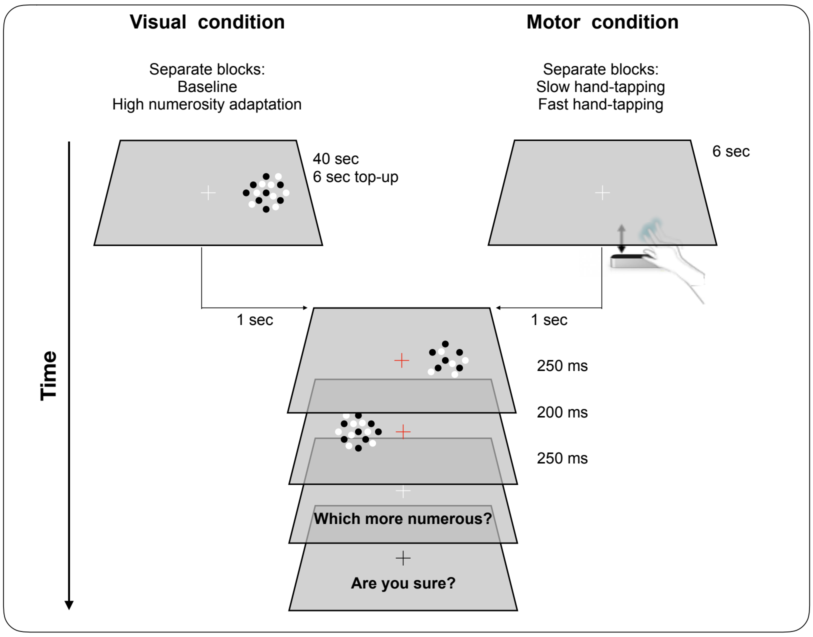


Figure 2.2. Stimuli and procedure. On each trial subjects were required to indicate which of two stimuli were more numerous, then report whether they were confident with their response (both responses 2AFC). In the visual adaptation condition, a dense dot array was displayed first for 40 s than for 6 s top-up periods at the test location before the discrimination task (top left). In the motor adaptation condition (top right), participants were required to tap their hand with index finger extended, for 6 s on the right side of the screen, with their hand concealed by the screen and without touching any surface to minimize sensory feedback. Subjects either tapped as fast as possible or slowly, at around 1 Hz (tested in separated sessions). In all conditions, reaction times between the offset of the reference and the numerosity response were measured, although participants were never requested to make any speeded response.

2. 2. 3 Data analysis

The proportion of trials where the test appeared more numerous than the probe was plotted against physical numerosity and fitted with cumulative Gaussian error functions. The median of the error functions estimates the point of subjective equality (PSE), and the difference in numerosity between the 50% and the 75% points gives the just notable difference (JND). The distributions of average confidence responses (1 for high, 0 for low) and of the mean of reaction-times were fitted with Gaussian distributions, and the peak of the fitted functions was taken as the point of maximum uncertainty or reaction-times.

$$P(N) = b + a \cdot exp\left(\frac{-(\bar{N}-N)^2}{2\sigma^2}\right) \qquad \text{eqn. 2}$$

Where *N* is numerosity, *P(N)* the proportion of confident responses – or the average reaction-time – at that numerosity, b and a constants, $\bar{N}$ the mean of the Gaussian and σ the standard deviation. When fitting data pooled over participants, all parameters were free to vary. When fitting individual participant data, b and σ were fixed to the values obtained for the aggregate data.

All analyses were performed both on the "aggregate participant", pooling all data from all participants, and also on individual participant data. Significance of the aggregate data was calculated by bootstrap sign test: 10,000 reiterations, with replacement.

Experimental procedures were approved by the local ethics committee (*Comitato Etico Pediatrico Regionale Azienda Ospedaliero-Universitaria Meyer, Florence*, Italy; protocol n. GR- 2013-02358262) and are in line with the declaration of Helsinki. All subjects gave written informed consent.

2.3 Results

2.3.1 Effects of adaptation on confidence and reaction times

We monitored decision confidence and reaction-times (in an un-speeded task) while participants made numerosity judgements after adaptation, either to dense visual patterns or to hand-tapping. The major results were obtained from analysis of the "aggregate observer", pooling data over all 12 participants (10 in the adaptation to hand-tapping). However, we also analysed individual data from all participants separately (**Figures A1&A2**) and, although the reduced data were necessarily more noisy, the group analysis gave essentially the same results as the aggregate. The results of the individual analyses are summarized in **Figure A4** and in **Table 2.1**.

Figure 2.3 shows the main results from the aggregate data. **Figures 2.3A and 2.3B** are psychometric functions, plotting the proportion of trials (for all participants) where the test was reported as more numerous than the reference, as a function of the numerosity of the test patch. Both data sets were well fit by cumulative Gaussian error functions, which were clearly displaced by adaptation, both by visual dot-patterns and hand-tapping. In the un-adapted condition (**Figure 2.3A**, blue symbols and curves), the psychometric function was centred at 17 dots, very near the actual reference of 16 dots. Visual adaptation to 60 dots clearly displaced the psychometric function rightwards, shifting the median (which estimates the PSE) to 22.7 dots, meaning that after adaptation the probe needed to be 33% more numerous than the reference to appear equal to it. A similar effect occurred for hand-tapping: slow tapping had little effect, with the PSE remaining at 15.9 (near the reference), while fast taping increased it to 18.1, again implying a decrease of apparent numerosity, in this case of 14%.

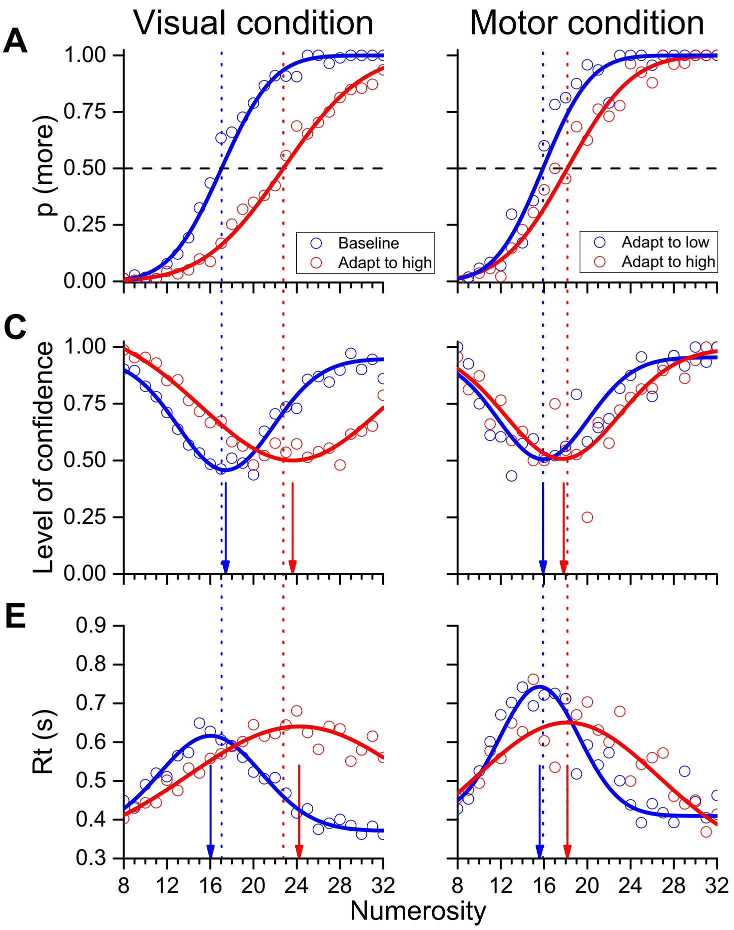


Figure 2.3. A-B: **Psychometric, confidence and reaction times functions.** Psychophysical functions showing proportion of trials in which the test was perceived more numerous than the reference, as a function of test numerosity. C-D: Confidence levels and mean reaction times (E-F) as a function of test numerosity, for visual and motor adaptation (left and right panels respectively). In all graphs, blue and red curves indicate baseline and high adaptation for visual adaptation (panels on left hand side) and slow or fast tapping in the motor experiment (on right hand side). The dashed lines show the PSEs and arrows the peaks of the best-fit gaussians to the confidence or reaction time distributions.

Both the confidence and mean reaction-time data were well fit by Gaussian functions ($R^2 > 0.75$ in all cases). The peaks of these functions (indicated by the arrows, and reported in **Table 2.1**), clearly also shift with adaptation, both to visual numerosity and hand-tapping. The shift is in the same direction as the shift in PSEs, tending to align peaks in confidence and reaction-times with the PSEs. These results on the aggregate observer are very similar to those obtained from analysis of individual participants (**Figure 2.4**).

Conditions			Aggregate		Individual Participants	
			Mean	sem	Mean	sem
Visual	Baseline	PSE	17.05	0.14	17.02	0.39
		Peak Conf.	17.42	0.18	17.53	0.54
		Peak RT	16.03	0.19	16.29	0.79
	Adapt to high	PSE	22.76	0.20	22.82	0.87
		Peak Conf.	23.63	0.34	23.16	0.88
		Peak RT	24.22	0.49	23.35	1.09
Tapping	Slow adaptation	PSE	15.89	0.25	15.79	0.29
		Peak Conf.	15.92	0.44	15.62	0.72
		Peak RT	15.58	0.27	15.64	0.47
	Fast adaptation	PSE	18.15	0.32	18.1	0.34
		Peak Conf.	17.67	0.40	16.82	0.52
		Peak RT	18.15	0.50	18.22	0.52
Control	Baseline	PSE	17.48	0.15	17.39	0.41
		Peak Conf.	17.07	0.29	17.57	0.74
		Peak RT	17.35	0.29	17.59	0.74
	Reward-low	PSE	15.776	0.15	15.734	0.56
		Peak Conf.	17.44	0.32	17.45	0.43
		Peak RT	16.30	0.20	16.86	0.44
	Reward-high	PSE	19.16	0.17	19,06	0.64
		Peak Conf.	17.98	0.22	17.47	0.84
		Peak RT	17.57	0.26	17,67	0.39

Table 2.1. **PSEs, minimal confidence and peak reaction times for all three experiments.** Data from the aggregate participant are shown on the middle column, average of individual participants on the right.

The blue and red histograms of **Figure 2.4(A-D)** show the results of bootstrapping (10,000 repetitions, sampling with replacement). On each repetition, estimates were made for PSE, point of *minimal confidence* and *maximal reaction-time*. It is clear from inspection that in all cases the distributions for the investigated conditions overlap very little, indicating that they are significantly different. Bootstrap sign test yielded significance levels of $p < 0.003$ in all cases. On adaptation to visual stimuli peaks in both the confidence (**Figure 2.4A**) and reaction-time (**Figure 2.4C**) were higher for the adapt-high condition than baseline in all 10,000 iterations ($p < 10^{-4}$). On adaptation to tapping, peaks in confidence (**Figure 2.4B**) were lower for the adapt-high than adapt-low condition on only 34 iteration ($p = 0.0034$), and for reaction-times (**Figure 2.4D**) only 20 times ($p = 0.002$) out of 10,000.

We then used the bootstrapped distributions to pit two plausible models against each other: 1) that the shifts in the psychometric functions result from a response strategy for uncertain trials (Morgan et al., 2011: illustrated in **Figure 2.1**); 2) that the change reflects adaptation-induced changes within sensory circuits. Model 1 predicts that the confidence and reaction-time distributions should not move with adaptation, so those for the adapt-high should be closer to PSE_{base} (or PSE_{low}) than to PSE_{high}. On the other hand, model 2 predicts that both peaks should follow the shifts in PSE, and therefore be closer to PSE_{high}. We tested this by bootstrap sign test, counting how many iterations were closer to PSE_{base} (or PSE_{low}) than PSE_{high}. We also bootstrapped the PSEs themselves on each iteration, to include their error in the calculation (the orange distribution in **Figure 2.4** shows the bootstrapped mid-points of the two PSEs). For visual adaptation, not a single iteration of either confidence or reaction-time peaks was closer to PSE_{base} than PSE_{high}, implying the likelihood for the first model is $p < 10^{-4}$. The tapping condition also showed a clear effect. For the confidence data, the likelihood of model 1 was $p = 0.05$, compared with $p = 0.95$ for model 2, giving a likelihood ratio of 19. Reaction-times were more significant, with likelihood of model 1 equal to 0.0064 compared with 0.9936 for model 2, 166 times less likely. All the bootstrapped sign tests provide strong evidence for model 2 for both types of adaptation, suggesting that the adaptation occurs within sensory rather than decision systems.

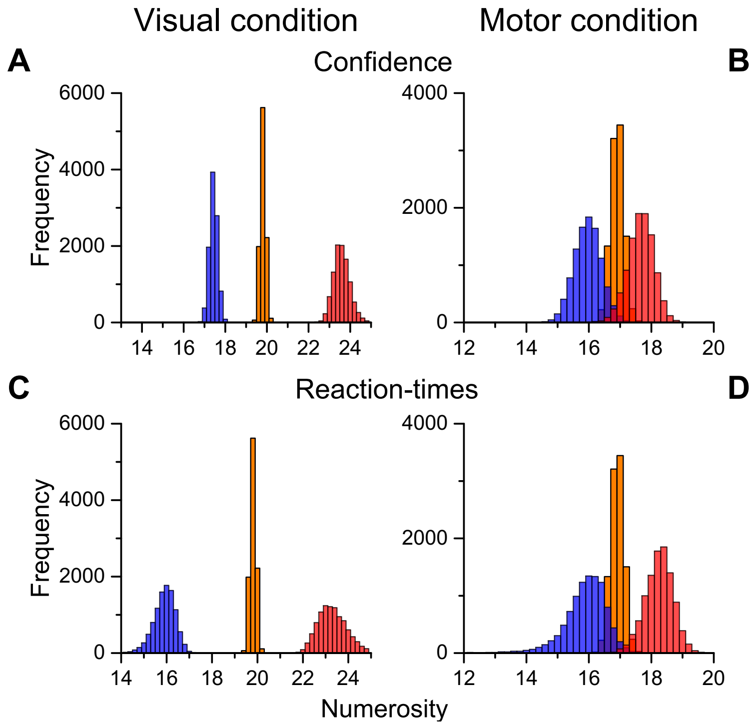


Figure 2.4. Frequency distributions of bootstrapped data of confidence and reaction-times. Frequency distributions of bootstraps for confidence (A-B) and reaction-times (C-D), for visual or motor adaptation experiment (left and right panels respectively). Data in blue represent visual baseline or slow tapping condition and red for high visual adaptation or fast motor tapping). Orange distributions show the bootstrapped mid-points between baseline (or slow) and adaptation (or fast tapping) PSEs.

To test the validity of the confidence ratings, we separated the data into high- and low-confidence trials and fitted psychometric functions separately for each, calculating the just noticeable difference (JND), from the standard deviation of the fit. Standard errors and significance were calculated by bootstrap. As there were 3 times as many trials judged confident than unconfident, the data for confident judgements were under-sampled during bootstrapping to match sample sizes. **Figures 2.5A and 2.5B** show JNDs for the high-confidence trials were significantly lower than that for low-confidence, by at least a factor of two ($p < 10^{-4}$ in all cases), consistent with the idea that subjective confidence reflects a genuine metacognitive ability which assesses the quality of sensory evidence (Mamassian, 2016).

We also correlated reaction-times against confidence (**Figures 2.5C and 2.5D**). Each point of **Figure 2.5C** comes from **Figures 2.3C and 2.3E**, and those from **Figure 2.5D** from **Figures 2.3D and 2.3F**. The correlation was strong, with $r = -0.87$ and -0.89 for the two adaptation types, accounting for more than 70% of the variance. This shows that the two measures covary together, consistent with their being driven by a common factor, most probably perceived stimulus strength.

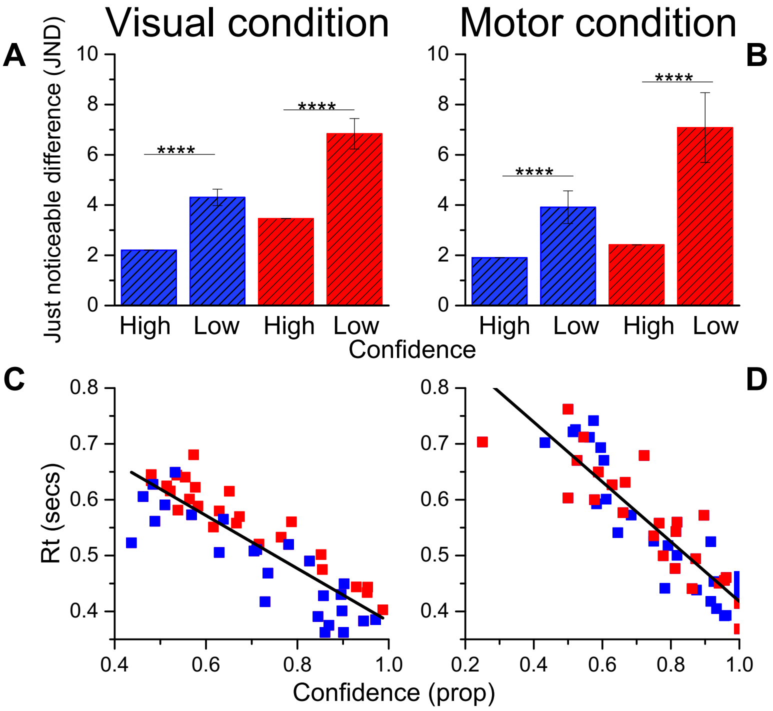


Figure 2.5. Just noticeable difference of confidence split-data and correlations between reaction-times and confidence in visual and motor conditions. Bar graphs show precision for numerosity discrimination in the high or low confidence trials. In blue, data for baseline (or slow tapping) and red data for adaptation to high (or fast tapping) for visual and motor adaptation. (C-D) Reaction-times (averaged over trials and subjects) as a function of confidence (averaged over trials and subjects) for the two adaptation conditions. Black lines represent the best-fitting linear regressions (C visual adaptation: R2 = 0.76; D motor adaptation: R2 = 0.79). Error bar represent ±1 s.e.m., *** $p < 0.0001$.

2.3.2 Control experiment: Effects of reward on confidence and reaction times

In order to show that confidence and reaction times do not necessarily change with PSE, we devised a control experiment where we manipulated rewards. Here there was no adaptation, but participants played a point-based game, with three types of reward regimes (in different blocks). In baseline blocks, they received 1 point for each correct response and lost 1 for every error (performing on average at 85% correct). In "reward-low" blocks, they received 2 points for correctly responding "less than", and lost 1 each error; and in "reward-high", 2 points for correctly responding "greater than", losing 1 for an error. This simple manipulation of rewards biased observers towards the double-reward response when uncertain, causing robust shifts in the PSE. **Figure 2.6A** shows the psychometric functions for the aggregate observer for the three conditions. The PSE for the standard condition was 17.5 (a constant bias of 1.5 from the physical equivalent of 16), while for the "reward-low" condition it was 15.8 (1.7 lower) and for "reward-high" was 19.1 (1.6 higher). Both cases are near the predictions of the ideal observer (which predicts a shift of 1.2 towards the rewarded side).

However, the shift in PSE was not accompanied by concomitant shifts in confidence: the minima in the gaussians are very similar for all three conditions (17.4, 17.1 & 18.0 for low, baseline and high). Similarly, the peak reaction times did not follow the PSEs, but again tended to cluster around the baseline PSE (16.3, 17.3 & 17.6). The histograms below the confidence and RT curves show the bootstrap analysis, similar to that of **Figure 2.4**. The bootstraps clearly overlap considerably. Again, we tested the two plausible models outlined for **Figure 2.4**, counting, for each condition, how many iterations were nearer to the PSE of that condition rather than to the PSE of the baseline (non-rewarded) condition. For the confidence measures the results were clear: the probabilities of model 2 (closer to the shifted PSE) being correct were $p=0.046$ for the reward-low condition, and $p=10^{-4}$ for the reward-high condition, 20 and 10,000 times less likely than model 1. The results for reaction times were similarly in favour of model 1, with probabilities for model 2 at $p<10^{-4}$ for the reward-low condition, and $p=0.012$ for the reward-high condition, infinite and 81 times less likely than model 1. Reaction times in this experiment may have been less reliable, because of variable slowing when integrating the reward "*prior*". Again, the results from the aggregate observer are very similar to those obtained from analysis of individual participants (see **Figure A3**).

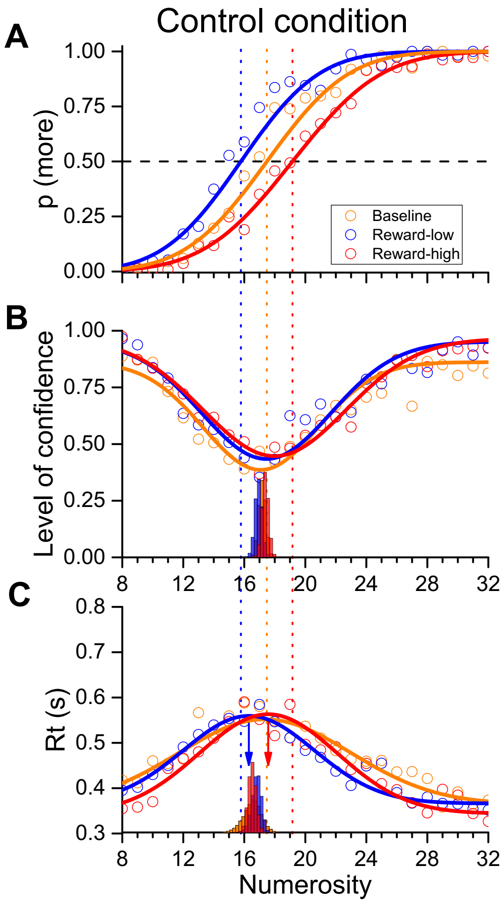


Figure 2.6. Psychometric, confidence and reaction-times functions of the control condition. (A) Psychophysical functions of proportion of trials when the test was seen as more numerous than the neutral probe, as a function of physical numerosity in the control condition (baseline in orange, leftward condition in blue and rightward condition in red). (B) Expressions of confidence, as a function of physical numerosity. (C) Mean reaction-times (in seconds) as a function of physical numerosity. The continuous dotted lines indicate the PSE of the psychophysical curves. The histograms below the confidence and reaction time fits represent the bootstrap analysis.

2.4 Discussion

The primary goal of this study was to probe the mechanisms of numerosity adaptation, to test whether adaptation affects sensory processing mechanisms directly, or indirectly via decision or response criteria. We argue that a change in sensory processing should result in a comparable change in minimum decision confidence and maximum reaction-times, which should shift to align with the point of subjective equality after adaptation, where the test and probe stimuli are, by definition, most similar perceptually. On the other hand, if the change in PSE results from a response bias, the peaks in confidence and reaction-times should not change with adaption (see **Figure 2.1**). Our results clearly support the claim that adaptation affects sensory processing directly. Two types of adaptation – to visual patterns and to hand-tapping – caused large shifts in PSEs, with concomitant shifts in peak confidence and reaction-times. In all cases, the sensory processing model was far more probable than that suggested by confidence-induced shifts in response criteria. On the other hand, when the PSEs were shifted by awarding rewards for specific responses, the shifts in PSE were not accompanied by shifts in confidence or RTs.

The results are interesting for several reasons. Firstly, there has been a long-standing debate about the nature of numerosity processing, particularly about whether it is sensed directly, or is a by-product of texture processing (Anobile, Cicchini, et al., 2016; Burr et al., 2018). One of the strongest lines of evidence that numerosity is distinct from texture density comes from adaptation studies, particularly cross-modal and cross-format adaptation (Arrighi et al., 2014): adapting to sequences of flashes or tones affects the perceived numerosity of dot arrays, difficult to ascribe to texture perception. The demonstration that adaptation to fast or slow hand-tapping changes the perceived numerosity of spatial arrays is even more fascinating, as it links perception and action, implicating common mechanisms for perceiving and reproducing numerosity (Anobile, Arrighi, et al., 2016, 2020).

However, paraphrasing Laplace (1812): “extraordinary claims require extraordinary evidence”. It is therefore reasonable to expect a rigorous demonstration that motor tapping affects the perception of numerosity directly, rather than merely biasing the decision or the response along the lines of **Figure**

2.1. The fact that all analyses show that both confidence and reaction-time peaks move to the adapted PSE strongly favours the hypothesis that adaptation causes changes at the sensory level. This has important ramifications for understanding the role of numerosity mechanisms in perception and action, relating well to the electrophysiological studies showing a clear selectivity for the number of self-produced actions in the area 5 of the superior parietal lobule of monkey (Sawamura et al., 2002, 2010).

The other more general result of this study is a method of validating adaptation and other effects of temporal and spatial dependency (such as serial dependence: see Cicchini et al., 2014, 2017; Fischer & Whitney, 2014; Fornaciai & Park, 2019). Adaptation is a fundamental tool in psychophysics, famously referred to as “the psychophysicist’s microelectrode” (Frisby, 1979). However, adaptation studies necessarily rely on subjective judgements, on participants reporting their subjective impressions. Most modern adaptation studies use two-alternative forced choice techniques that ask participants to compare the adapted test to a probe, yielding psychometric functions from which the point of subjective equality can be titrated. However, unlike other forced-choice tasks (such as measurement of contrast sensitivity), there is no right or wrong answer: just a subjective judgment that stimulus A was larger, brighter or more numerous than stimulus B. Over a considerable range around the point of subjective equality, judgments are difficult, but participants must respond, guessing if unsure. It requires only a slight tendency to respond stereotypically in one direction when unsure to shift the curves, robustly changing the PSE, without changing the slope of the function (Morgan et al., 2011). It therefore becomes important to have objective corroborative evidence that the point of subjective equality really does reflect sensory changes rather than response biases. Gallagher et al. (2019) suggested that minima in response criteria could provide useful corroboration, and demonstrated that they can do so for motion adaption (and also for serial dependence). We build on their idea, showing that even with a far more subtle forms of adaptation elicited by hand-tapping, the minima in confidence follow the changes in PSE.

We point out that we are testing a specific model of how decision criteria may affect PSEs; that a small tendency of response bias could affect trials of low confidence, causing reliable shifts in PSE (Morgan et al., 2011). With this particular model, as confidence is driving the response, this is unlikely to shift with the response PSE. However, other, more complex models of perceptual decisions (Maniscalco & Lau, 2014; Ratcliff & McKoon, 2008) may predict that confidence and RT do change with changes in PSE. Indeed, with these classes of models it is often difficult to distinguish experimentally between sensory and perceptual decision effects (Pelli, 1985). We therefore designed a realistic experiment that manipulated PSEs at the decisional level, by rewarding correct responses in a specific direction (high or low). This produced robust changes in responses, shifting the PSE as expected, as participants sought to optimize gains; however, the shifts in PSE were not accompanied by concomitant changes in confidence, nor in RTs. This is a clear existence proof that at least some types of manipulation on decisions are not paralleled by shifts in confidence, which may therefore be a signature of sensory changes. Gallagher et al. (2019) performed a similar experiment, instructing participants specifically to respond “left” or “right” when confidence is low, and also showed that this manipulation does not shift the point of minimal confidence. However, our task was more natural, in that we gave no instructions to participants on how to respond, nor that they should take confidence into account. It was a natural task with greater risks on one side than on the other (like those pioneered by Trommershäuser and colleagues (2005)) which human participants soon learn to optimize. Yet this very natural and spontaneous task, which shifted PSEs smoothly, caused no similar shifts in confidence or RTs.

In general, reaction-times provided more robust data than confidence for the sensory shifts in PSE. Reaction times could have several advantages to confidence measures. Firstly, they are objective and come at no extra cost, automatically encoded in the timestamps of the stimuli and responses, without having to ask participants to make a second response. Nor was it necessary to ask for a speeded response; we simply relied on the tendency of participants to respond reasonably quickly in order to finish the session as soon as possible. For the adaptation experiments, reaction-times proved to be more informative than confidence, in all cases providing stronger evidence for a shift in their peak. For example, for the aggregate data for adaptation to tapping, the $\text{Log}_{10}\text{BF}_{12}$ was 1.26

for confidence, compared with 2.22 for reaction-time data. For the analysis of individual data (where there are far fewer trials, hence more noisy estimates) the $\mathrm{Log}_{10}\mathrm{BF}_{12}$ for confidence was 1.14 compared with 2.46 for reaction-times. In all cases the $\log_{10}$-Bayes factors were greater than 1, considered *strong* evidence, but the reaction-time data gave $\log_{10}\mathrm{BF} > 2$, considered *decisive* (Jeffreys, 1998). There is considerable evidence showing that reaction times vary monotonically with signal strength (Piéron, 1914), and should therefore be maximal at the point of least difference in the signals. Combined with the ease with which reaction-time data can be collected, with no additional load on participants, it would appear to be the preferred method.

To summarize, we present a new technique for investigating the mechanisms of numerosity adaptation and sensory adaptation in general. By simultaneously measuring subjective confidence and more importantly – reaction-times, we demonstrate that adaptation to numerosity, either by observing visual stimuli of high numerosity or by subjects tapping in a particular region occurs at a sensory level, before stages of perceptual decision. Adaptation affects not only perceived numerosity, but also subjective confidence and reaction times, showing that they are a consequence of sensory adaptation, rather than the cause for the shift in the psychometric functions.

Chapter 3

“Groupitizing”: a strategy for numerosity estimation

Chapter 3

3.1 Introduction

Recently, Starkey and McCandliss (2014) suggested that subitizing mechanisms may also come into play for higher numerosities, a process they term "*groupitizing*". This is very much like George Miller's well-known notion of “chunking”, where complex sets of information such as long telephone numbers can be more easily recalled if parsed into three or four smaller “chunks”. Starkey et al. (2014) measured counting speed of spatially clustered arrays in school-age children, and found that clustering, or grouping, increased performance. Crucially, both the number of clusters and the number of elements within each cluster was limited to the subitizing range (e.g. 7 = 2+2+3). Interestingly, the grouping advantage increased with age and correlated with arithmetic abilities, with more math-skilled children showing stronger groupitizing effects. More recently, it has been reported that grouping by color can also decrease reaction times in adults (Ciccione & Dehaene, 2020). Overall these studies suggest that serial counting without time constraints may be not a "pure" and direct measure of ANS precision, but could be tempered by arithmetical strategies, such as grouping, which involves processes such as parse-and-add.

In this study, we ask whether grouping items by spatial proximity or color not only increases enumeration speed but also increases precision (measured as Coefficient of variation). We also investigated whether this grouping phenomenon is a general property of numerosity perception, applying to temporal sequences as well as spatial arrays. The results suggest that groupitizing occurs for estimation of both temporal and spatial dimensions of numerosity. We also observed a robust inter-individual variability in the magnitude of grouping-based improvement, with participants who were less precise in estimating numerosity in ungrouped arrays benefiting more from the groupitizing. This suggests that some participants may take advantage of intrinsic grouping in ungrouped arrays to increase their performance, therefore benefit less from the explicit experimentally induced grouping.

3.2 Methods

Sixteen young adults (mean age=26, standard deviation=3.2, range= 23-36) participated in this study (12 male, 4 female, 13 participants were master's students in psychology, 2 were grad-students and 1 a post-doc in neuroscience). All participants had normal or corrected-to-normal vision. All completed all tasks except one, who was unavailable to perform the two sequential numerosity conditions.

Stimuli were created with Psychophysics toolbox for Matlab and displayed on a 60 Hz – 15" screen monitor (MacBook Pro) placed at viewing distance of 57 cm. Subjects were tested in a quite, dimly light room. The experimental procedures were approved by the local ethic committee (*Comitato Etico Pediatrico Regionale* — *Azienda Ospedaliero-Universitaria Meyer* — *Firenze FI*). The research was performed in accordance with Declaration of Helsinki and informed consent was obtained from all participants prior to each experiment.

Each trial started with a central fixation point that remained on screen for the entire experiment. After 500 ms a stimulus was displayed, followed by a blank screen. Participants estimated verbally the numerosity of the squares-array or square-sequence (in separate sessions with order pseudorandomized between subjects **Figure 3.1C and 3.1D**).

The experimenter hit the spacebar when the participant responded (used to calculate reaction times), then entered the response on the numeric keypad, which initiated the following trial. Response time was measured from the stimulus offset to the beginning of vocalization. Participants were asked to respond as soon as possible, but also to concentrate on accuracy. Each condition was tested in separate blocks, and participants were never explicitly informed about the grouping cues.

Numerosity levels ranged from 4 to 16 (grain of 1, resulting in 13 numerosity levels). In the structured conditions, each numerosity was organized into clusters (between 2 and 4), each containing a variable number items (between 2 and 6), resulting in the following configurations: 2, 2 – 2, 2, 1 – 3, 3 – 2, 2, 2 – 3, 3, 1 – 3, 3, 2 – 2, 2, 2, 2 – 4, 4 – 4, 3, 2 – 4, 4, 1 – 3, 3, 3 – 3, 3, 3, 1 – 4, 4, 2, 1 – 3, 3, 3, 2 – 3, 3, 3, 3 – 4, 4, 4 – 4, 4, 3, 3 – 4, 4, 4, 3 – 5, 5, 3 – 4, 4, 4, 4 – 5, 5,

6. As numerosities 4 and 16 were not analyzed (see data analyses), each grouped pattern comprised a minimum of 2 and a maximum of 4 clusters. All clusters except one (13 = 5, 5, 3) contained from 1 to 4 elements. On each trial, a given numerosity and configuration pattern were randomly selected. Each participant completed about 150 trials for each of the six conditions (around 14000 trials in total).

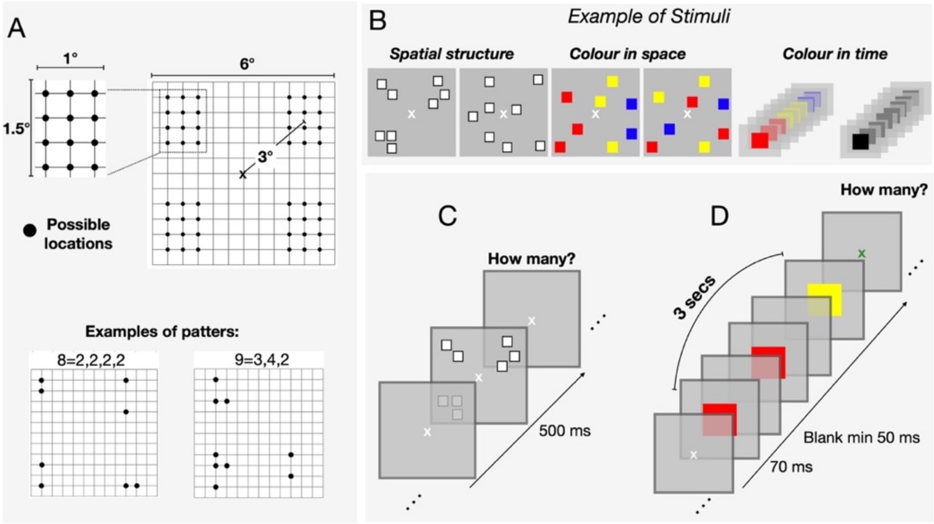


Figure 3.1. Stimuli and procedure. A) Illustration about how stimulus position was defined in the grouping conditions (upper panel) with example configurations for numerosities 8 and 9 (lower panel). B) Examples of stimuli arrangement in the various conditions, when grouping was defined by spatial proximity, color or temporal proximity, together with related ungrouped conditions (on the left-hand side). C-D) Example of the time course for the spatial (C) and temporal (D) version of the experiment (tested in separate sessions). In the spatial numerosity conditions (C), a central patch of squares was presented for 500 ms. In the sequential numerosity condition (D) a series of squares was centrally presented. Participants were asked to verbally report the perceived numerosity. Stimuli are not depicted to scale.

3.2.1 Stimuli

<u>3.2.1.1 Spatial arrays</u>

Stimuli were arrays of squares (0.4° × 0.4°) displayed for 500 ms on each trial. Squares could not overlap and were constrained to fall within a 6° × 6° virtual square area. In the conditions where spatial structure was manipulated, the individual items were white squares within black borders (so luminance was not a cue to number). In the unstructured conditions, the position of each square was randomly selected from 154 possible positions (within the stimulus area), being the centers of equally spread sectors within the 6.5°×6.5° area (each grid 0.5°×0.5°). For the spatially grouped condition, stimuli were arranged in 4 possible groups of 12 possible positions (see **Figure 3.1A**). Each group (spanning over a max area of 1.5 × 1 deg) was located in one quadrant and centered at 3° from the central fixation point. Each group was first randomly assigned to one quadrant (between 1 and 4), then the individual items positions was randomly selected between one of the 12 in the selected quadrant. Within each quadrant, the maximum center-to-center distance between each element was 2° and the minimum was 0.5°.

In the conditions where groups were defined by color, individual items could be red, green, blue or yellow (RGB: 255 0 0; 0 255 0; 0 0 255; 255 255 0). Color was assigned from left to right, so that similar colors appeared in vertical rows. For example, in the 3, 3, 2 condition depicted in **Figure 3.1B** squares were first randomly located, then the first three squares (from the left border) were colored red, the next three yellow and the remaining two blue (colors randomly chosen for each group). In the unstructured color condition, positions were assigned with the same logic, but with colors assigned at random.

3.2.1.2 Temporal sequences

Stimuli were streams of 3° × 3° squares each presented at screen center for 70 ms, for a total trial duration of 3 secs (**Figure 3.1D**). The end of each trial was signaled by color change of the central fixation point, from white to green. Sequences were spaced pseudo-randomly: on every trial, a given number of impulses (chosen at random) were evenly spread within the 3-second sequence duration; then the timing of each impulse was randomly jittered by either ±0, ±20 or ±40 ms to create a pseudorandom sequence of impulses with a minimum ISI between consecutive flashes of 50 ms. In the ungrouped condition all stimuli were black, while in the grouped condition they were grouped by color: each flash within a group could be red, green, blue or yellow (color coordinates as before), with group color randomly assigned. For example, in the 3, 3, 2 condition depicted in **Figure 3.1B**, the first three flashes were colored red, the following were yellow and the remaining two blue.

3.2.2 Data analysis.

Since participants were explicitly informed about the numerical range (4-16), we eliminated the two extreme numerosities from the analyses. We controlled for response outliers by eliminating trials with RTs longer than 3 standard deviations from the average response time, calculated separately for each numerosity level and participant.

For each participant, we calculated for each numerosity the average perceived numerosity, the standard deviation of the responses and the median response time. Precision was measured by normalizing the standard deviation by the physical numerosity yielding the Coefficient of variation (CV), a dimensionless index of precision that allows comparison and averaging of performance across numerosities.

$$Coefficient\ of\ Variation = \frac{\sigma_i}{N_i} \qquad \text{eqn 1}$$

Where N_i is the analyzed numerosity and σ_i the standard deviation of responses to numerosity i. Improvement (I) by grouping was measured by a normalized index yielding the proportion improvement:

$$I = \frac{CV_R - CV_G}{CV_R} \qquad \text{eqn 2}$$

Where CV_R and CV_G are the Coefficient of variation for the ungrouped and grouped conditions.

Data were analyzed by Repeated Measures ANOVAs, and effect sizes were reported as η^2, using JASP and Matlab.

3.3 Results

We asked participants to estimate the numerosity of briefly presented visual impulses, presented either in simultaneous spatial arrays or temporal sequences. For both conditions (tested in separate sessions), we investigated the effects of task-irrelevant grouping cues on numerosity estimation precision and speed. Grouping manipulations mainly followed the formal definition of Starkey and McCandliss (2014) with both the number of groups and the number of items/events within each group falling within the subitizing range: 2, 3 or 4 groups each containing 1, 2, 3 or 4 items/events.

3.3.1 Effect of grouping on perceived numerosity

We first evaluated the effect of grouping on the accuracy of estimation of perceived numerosity. **Figure 3.2** shows averaged responses as a function of physical numerosity. To statistically test differences across conditions, we ran Repeated measure ANOVAs (one for each numerosity format: simultaneous and sequential) with numerosity (11 levels, from N5 to N15) and grouping condition (4 or 2 levels for simultaneous and sequential numerosity respectively) as within subject factors. For both numerosity formats, the main effect of numerosity was obviously significant (simultaneous: $F_{(10,150)}= 834.289$, $p<0.001$, $\eta^2= 0.982$; sequential: $F_{(10,140)}= 282.289$, $p<0.001$, $\eta^2= 0.953$), but there was no significant effect of "condition" (simultaneous: $F_{(3,45)}= 1.285$, $p= 0.29$, $\eta^2= 0.08$; sequential: $F_{(1,14)}= 0.281$, $p= 0.60$, $\eta^2= 0.02$) and the condition-by-numerosity interactions were insignificant (simultaneous: $F_{(30,450)}= 0.742$, $p= 0.84$, $\eta^2= 0.047$; sequential: $F_{(10,140)}= 0.311$, $p= 0.97$, $\eta^2= 0.022$). Overall, these results clearly indicate that grouping did not significantly affect average perceived numerosity.

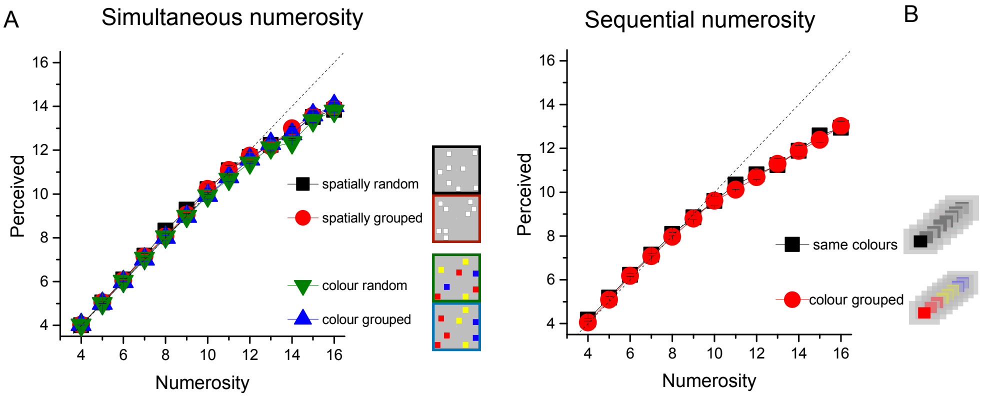


Figure 3.2. Perceived numerosity. Average perceived numerosity for spatial (A) and temporal (B) numerosity tasks, averaged across participants.

3.3.2 Grouping and sensory precision

Having established that grouping did not change average perceived numerosity (accuracy), we investigated its effect on sensory precision, indexed by Coefficient of variation (eqn. 1). This is a classical psychophysical parameter and, in the case of numerosity, is believed to reflect the sensory noise associated with the estimation process: higher values reflect less precision in the estimates and thus more sensory noise. **Figure 3.3** shows Coefficient of variations averaged across numerosities and participants for the ungrouped and grouped conditions, for estimations of spatial (A) and temporal (B) numerosity.

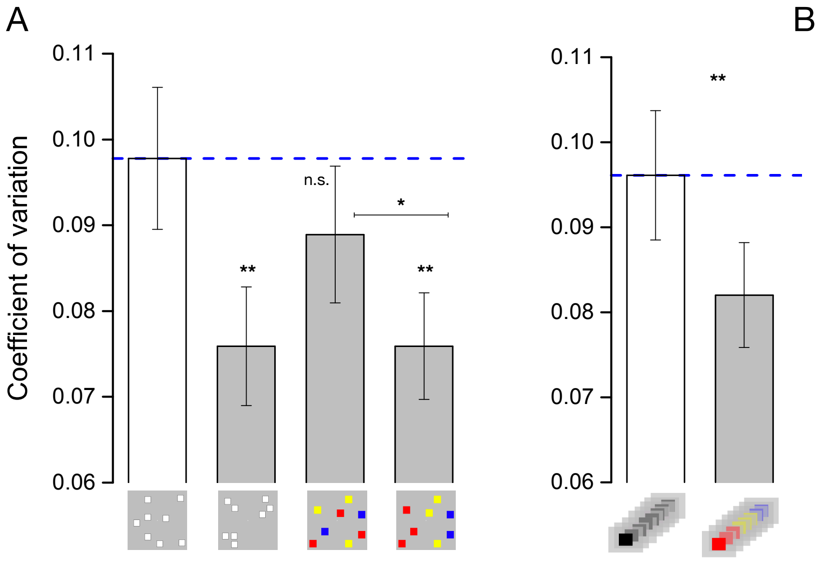


Figure 3.3. Sensory precision. Average Coefficient of variation for the simultaneous (A) and sequential (B) numerosity tasks. (A) Bar graph of average Coefficient of variation in the four spatial conditions: the stimuli could be randomly presented, grouped by spatial proximity (all items were white squares), randomly presented but coloured and grouped by color. (B) Bar graph of average Coefficient of variation in the two sequential conditions: the sequence of flash could be randomly presented (all flash were black) and grouped by color. Error bar represent ±1 s.e.m. $^{**}p \leq 0.01$ $^{*}p < 0.05$

For spatial presentations, Coefficient of variation was highest for the non-grouped condition, higher than all the grouped conditions. Repeated measures ANOVA with numerosity (11 levels) and condition (4 levels) revealed a significant main effect of condition ($F_{(3,45)} = 4.9$, $p=0.005$, $\eta^2 = 0.247$), with grouping decreasing Coefficient of variation compared to the spatially ungrouped condition (**Figure 3.3A**). The effect of numerosity was also significant ($F_{(10,150)} = 4.921$, $p<0.001$, $\eta^2 = 0.634$), suggesting that Coefficient of variations are not constant with numerosity, while the interaction was not ($F_{(30,450)} = 1.365$, $p = 0.097$, $\eta^2 = 0.08$), suggesting that the overall effect of grouping was constant across numerosity levels.

To assess the effect of grouping separately for each condition, we then ran a series of repeated measures ANOVAs against the spatially ungrouped stimuli condition. The results revealed that grouping by spatial structure ($F_{(1,15)} = 9.38$, $p =$

0.008) and by color gradient in space ($F_{(1,15)}$= 13.908, p= 0.002) both induced a significant reduction of Coefficient of variation, and both had a quite large effect (spatial structure 22%, η^2= 0.43, color gradient in space 22%, η^2=0.48). Grouping by color without spatial gradient did not produce a significant reduction in Coefficient of variation (9%, $F_{(1,15)}$= 2.264, p= 0.15, η^2= 0.13). The ANOVA comparing the two color conditions (with and without a spatial gradient) revealed that grouping by color with a gradient in space produced a significant reduction in Coefficient of variation compared to color alone (14% reduction in WF, $F_{(1,15)}$= 5.165, p= 0.038, η^2= 0.256). The interaction between numerosity-by-condition was never significant (p>0.05) in any condition comparison (**Figure 3.4A-D**), suggesting the effect was comparable across numerosity levels.

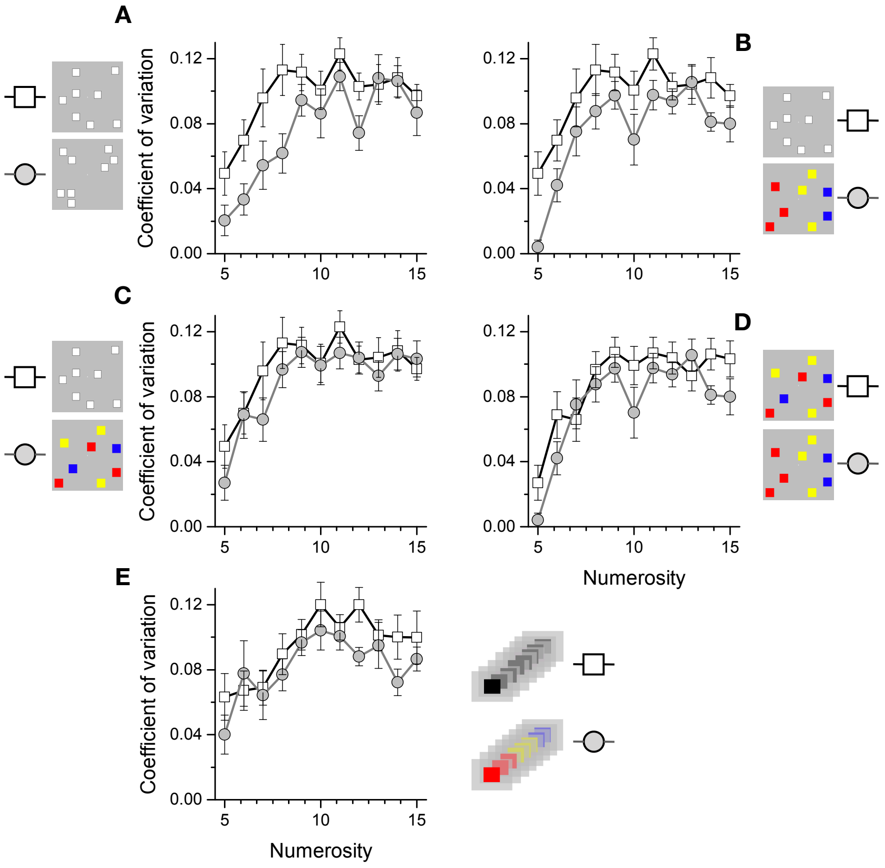


Figure 3.4. Sensory precision across numerosity levels. Average Coefficient of variation as a function of numerosity levels for all different experimental conditions. A-C: Performance in the spatially ungrouped condition (open squares) against grouping by spatial proximity (A, gray circles), color gradient in space (B,

gray circles) or only color (C, gray circles). D) Coefficient of variations for the two-color conditions, ungrouped or grouped by color. E) Coefficient of variation for the sequential presentation when stimuli were shared the same color (black) or similar colored items were presented temporally close to each other. Error bar represent ±1 s.e.m.

Figure 3.3B shows the effects of grouping on sequential numerosity. Here, grouping was encouraged with sequences of same-colored flashes within the sequence. Again, grouping yielded a clear increase in precision compared to the ungrouped condition, with a Coefficient of variation reduction of about 15% ($F_{(1,14)}$= 11.683, p= 0.004, η^2= 0.455). Once again, the numerosity-by-condition interaction was not significant (p>0.05).

3.3.3 Grouping and response times

Like previous studies in the literature (Ciccione & Dehaene, 2020; G. S. Starkey & McCandliss, 2014), we also investigated the effect of grouping in term of response speed (**Figure 3.5**). Reaction times were around 2 seconds for all experiments with spatial arrays (**Figure 3.5A**), and around 1.2 secs for the temporal sequences (**Figure 3.5B**).

Repeated measure ANOVA with numerosity (11 levels) and condition (4 levels) as factors did not reveal a significant effect of spatial grouping condition ($F_{(3,45)}$ =1.008, p=0.40, η^2= 0.06). However, separate repeated measure ANOVAs against spatially ungrouped stimuli revealed that grouping by spatial structure significantly reduced RTs from 2.02±0.26 to 1.86±0.22 secs, an effect of 8% (F($_{1,15}$)= 4.612, p= 0.048, η^2= 0.235, for all the other ANOVAs min p=0.25). There was a significant reduction of response time induced by grouping of temporal sequences (RT unstructured= 1.27±0.083, RT grouped= 1.14±0.072, an effect of 10%: $F_{(1,14)}$=8.861, p=0.01, η^2=0.388). Again, the effect of numerosity was statistically significant ($F_{(10,140)}$= 10.13, p<0.001, η^2= 0.42) but not the numerosity-by-condition interaction ($F_{(10,140)}$= 0.924, p= 0.513, η^2= 0.062). Finally, all

ANOVAs revealed a statistically significant effect of numerosity (reaction times increased with set size, all $p<0.001$), but no numerosity-by-condition interactions ($p>0.05$).

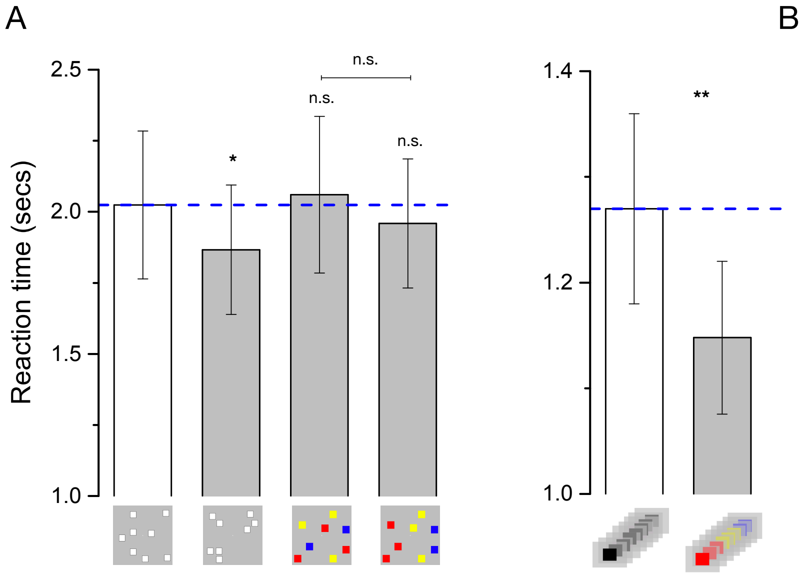


Figure 3.5. Reaction times. Average Reaction times for the various experimental conditions for simultaneous A) and sequential numerosity (B) formats. Error bar represent ±1 s.e.m. ***$p<0.001$, **$p<0.01$, *$p<0.05$

3.3.4 Interindividual differences in grouping advantage

The results so far show that grouping stimuli into easily separable, subitizable chunks yielded more precise estimates than with ungrouped patterns. The effect is robust, but there is also considerable interindividual variability. Here we asked whether the magnitude of improvement may be related to the baseline sensory precision. It is feasible that some participants always use grouping strategies to some extent, taking advantage of the intrinsic clustering of ungrouped patterns. If this were the case, we would expect these participants to benefit less from explicit grouping, as they were already using this strategy. That is to say, participants with the highest Coefficient of variations measured in the ungrouped conditions should benefit the most from the explicit grouping.

To test this notion, we correlated the magnitude of the grouping advantage (the normalized improvement by grouping (eqn. 2) against the baseline Coefficient of variation (**Figure 3.6**)). If grouping was to reduce all Coefficient of variations proportionally (multiplicatively), the correlation should be zero. If the effects were additive, then the correlation would be negative (proportionally greater for the lower Coefficient of variations). However, if those who had the highest Coefficient of variations profited proportionally more than those with lower Coefficient of variations, the correlations should be positive.

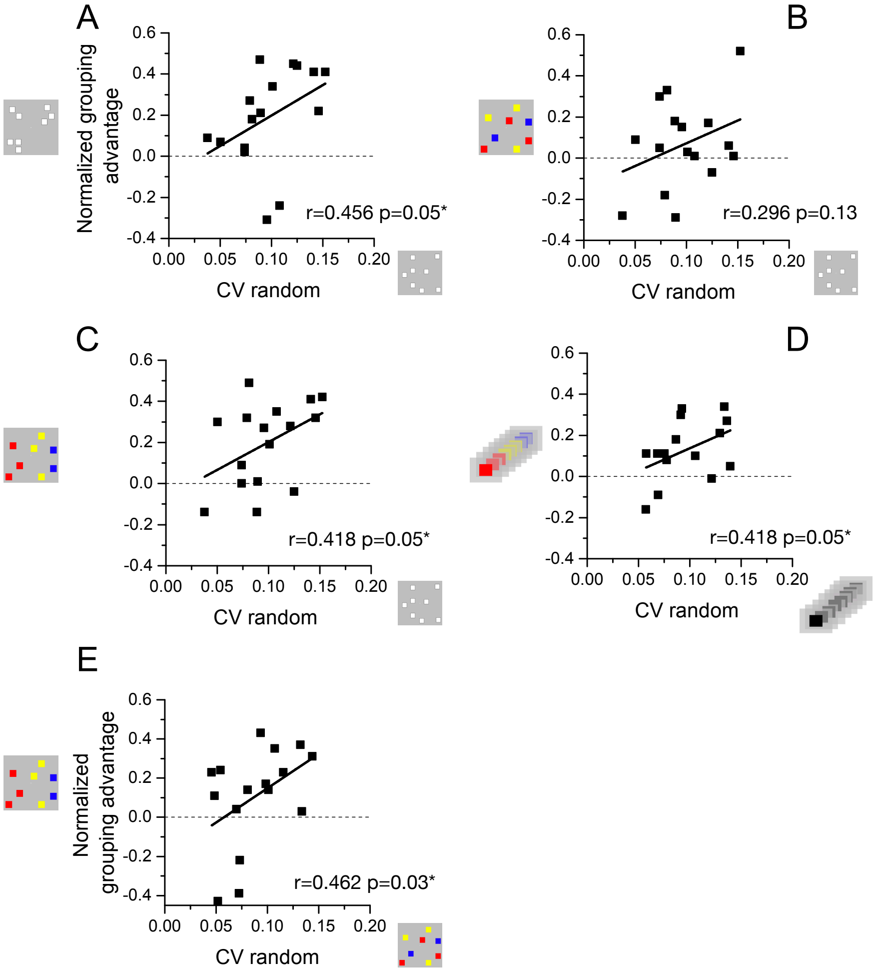


Figure 3.6. Individual differences. Correlations between grouping effects (normalized improvement by grouping) on estimation precision and Coefficient of variation in the different experimental conditions (A: spatial grouping, B, C & E: colour groping on spatial numerosity, D: colour groping on sequential numerosity). Positive values indicate a reduction of Coefficient of variation induced by grouping. Positive correlations (Pearson r) indicate that participants with worse sensory

precision in the unstructured conditions (abscissa) gained more from grouping. Lines are best linear fit, one tailed p-values. *p≤0.05

For the four conditions that yielded a significant grouping effect – spatial grouping, color clustering (with and without spatial grouping) and temporal color clustering – the correlation was significantly positive ($p < 0.05$, one-tailed test). On the other hand, the condition in which grouping did not yield a significant advantage on numerosity precision (ungrouped space Vs ungrouped color in space), showed no significant advantage ($p = 0.10$).

3.4 Discussion

This study shows that using color, or spatial or temporal proximity to group items together robustly improves the precision of numerosity estimation, by up to 20%.

The magnitude of the advantage for grouping did not vary with numerosity, over the range tested, from 5 to 15 (**Figure 3.4**). That is interesting, as one may have expected proportionally greater effects for the larger numbers. But perhaps there was also a greater cost in subitizing and doing addition with larger numbers, so the net proportional gain was similar. We selected our number range to be comfortably inside the range where numbers are thought to be estimated directly, rather than via texture-density mechanisms (Anobile, Cicchini, et al., 2016; Burr et al., 2018). It would be interesting to test much higher numerosities and densities, to see if grouping can also aid in judgments of texture density. It would seem unlikely if based on subitizing, as subitizing is limited to about 4, but worth verifying.

We also found smaller and less robust advantages in reaction times, confirming previous studies (Ciccione & Dehaene, 2020; G. S. Starkey & McCandliss, 2014). We found that grouping by spatial structure slightly reduced reaction times relative to the spatial ungrouped condition, by about 8%. However, RTs in the spatial gradient color condition were not significantly different from the ungrouped color condition. One possibility for this discrepancy may be that grouping by the spatial dimension is more salient compared with grouping by color. Alternatively, colored items may induce a strong tendency to automatically group the stimuli, even when randomly scattered spatially. This idea is supported by the lower CVs in the ungrouped condition with coloured stimuli compared to those measured with achromatic stimuli. Not surprisingly, this statistically insignificant trend was not evident in the RTs, in line with the fact that in the present study RTs have proven to be less robust in detecting grouping effects than CVs.

Previous research has shown that grouping, or *groupitizing*, speeds up serial counting (G. S. Starkey & McCandliss, 2014), but this does not help preschoolers. Furthermore, the grouping advantage correlated positively with arithmetical abilities in school-age children, suggesting that grouping relies, at least to some extent, on formal arithmetical knowledge. Thus, grouping may reflect an implicit math strategy of numerosity perception, like "parse the scene into subitizable groups then sum the subitized estimates". That grouping not only speeds counting but also lowers numerosity estimation thresholds has broad implications. Precision in numerosity estimation and discrimination are predictive of child math abilities (Anobile, Arrighi, et al., 2018; Halberda et al., 2008), and are both impaired in dyscalculia (Mazzocco et al., 2011; Piazza, 2010). These results have been interpreted as a link between the perceptual ability to estimate numerosity and the cognitive ability to learn math (Butterworth, 2019; Piazza, 2010). However, if grouping strategies are spontaneously used by some participants, such as those with more spontaneous arithmetical skills, it could be this that mediates the link between numerosity and math proficiency. Use of grouping information, either intrinsic or explicitly introduced, requires some basic math skills, such as rapid addition of the numerosities of the sub-groups. It is likely that participants who opt for this strategy – rather than a global appraisal of the whole pattern – would be those with the greater math skills. This would have important implications for understanding the link between measures of numerosity sensitivity and math.

In the present study, participants who were more precise in the ungrouped condition benefited proportionately less from grouped configurations than those with higher thresholds. One plausible explanation for this is that those with lower thresholds use grouping strategies even with the ungrouped patterns, taking advantage of intrinsic grouping in randomness. These people may benefit less from the explicit grouping imposed by spatial or temporal proximity, and therefore show less improvement. This possibility is interesting, with implications about different individual styles in numerosity perception, well worth pursuing further.

The correlation between numerosity precision and math skills is interesting. While thresholds for estimating numerosities at moderate, uncrowded densities predict well math performance (Anobile et al., 2013; Halberda et al., 2008; Piazza, 2010), numerosity discriminations at high densities (Anobile, Castaldi, et al., 2016) do not; nor does subitizing (Anobile, Arrighi, et al., 2019). Furthermore, thresholds for temporal sequences do not predict math performance (Anobile, Arrighi, et al., 2018), despite the clear evidence for a generalized number system encompassing space and time (Anobile, Arrighi, et al., 2016; Arrighi et al., 2014; Burr et al., 2018). All this suggests that some aspect of estimation of numerosity at low densities is related to math. A clear candidate mechanism could be "groupitizing", the use of strategic grouping to parse arrays into subitizable chunks. As mentioned above, this strategy requires some basic arithmetical skills: simple but rapid addition. It is reasonable to suppose that this skill does not help in the subitizing range, where arrays are already subitizable without further parsing, so that is not predictive of math. Similarly, for high numerosities the parsing strategy would not be effective, as only a limited number of subitizable sub-sets can be counted. Why estimation of temporal numerosity sequences does not correlate with math is less clear, as the present results show that a grouping strategy is possible with temporal sequences, and that those who benefit most from the grouping cues are those with highest thresholds. Perhaps the fact that temporal sequences are necessarily one-dimensional makes it harder to spontaneously group into sub-sets, particularly for young children. Also, in previous studies the presentations were constrained to be quite rhythmic, which does not lend to spontaneous parsing into groups. And perhaps phenomena such as "entrainment" tend to make the sequences even more rhythmical, and hard to group (Jones & Mcauley, 2005; McAuley & Jones, 2003). Again, this idea bears further investigation, particularly with children.

To conclude, the current study demonstrated that use of grouping strategies can aid considerably in the estimation of numerosity. The strategy may be related to mathematical abilities, and understanding it better could be of considerable importance in understanding the link between estimating numerosity and formal math skills.

Chapter 4

Grouping strategies in number estimation extend the subitizing range

Chapter 4

4.1 Introduction

Humans can generally count or estimate the number of objects in a scene quite easily, yet the perceptual mechanisms and the cognitive strategies underlying this ability are still little understood. Numerical judgments are extremely fast and virtually errorless up to four items, while they become slower or more approximate for larger numerosities (Atkinson et al., 1976; Jevons, 1871; Kaufman et al., 1949). This behavior suggests the existence of two independent systems for perception of very small and larger numerosities, the subitizing and the Approximate Number System (ANS) respectively (Dehaene, 2011).

Interestingly, counting speed of larger numerosities also increases considerably if stimuli are grouped into smaller clusters (Beckwith & Restle, 1966; Wender & Rothkegel, 2000), a phenomenon that has been termed *groupitizing* (G. S. Starkey & McCandliss, 2014). Counting is particularly fast when the number of clusters and the number of items included in each cluster is very low (e.g. 8= 4+4), falling within the subitizing range (G. S. Starkey & McCandliss, 2014). Two recent studies have generalized the groupitizing effect to non-spatial grouping cues, different numerosity tasks and formats. Ciccione and Dehaene (2020) showed a groupitizing advantage only when items were divided into clusters of the same number of items, irrespective whether the items were grouped spatially or by color alone. Anobile et al. (2020) went on to show that groupitizing can also boost sensory precision measured with an approximate numerosity estimation task, both for spatial arrays and temporal sequences. Starkey and McCandliss (2014) noticed that school-age children with higher arithmetical abilities took most advantage of groupitizing cues, while there was no groupitizing effect in preschoolers, suggesting that the ability to groupitize may reflect the use of arithmetical strategies (e.g. divide-and-sum).

A reasonable conclusion from these studies is that groupitizing arises from two independent factors: the ability to subitize small groups parsed from the larger set, and the ability to combine the group estimates through mental calculation. The first aspect implies that participants may recruit the subitizing system to estimate numerosities higher than the normal 4-item limit. This strategy would require considerable cross-talk between subitizing and ANS, usually considered to be independent systems. However, there is some evidence for interconnection between the systems. Under dual task conditions, sensory thresholds for estimating numerosities in the subitizing range become comparable to those measured in the estimation range, suggesting that the estimation system works even within the subitizing range, but performance for low numbers normally augmented by the automatic deployment of visuo-spatial attentional resources (Burr et al., 2010, 2011; Pomè, Anobile, Cicchini, Scabia, et al., 2019). The heavy reliance of subitizing on attention may therefore constitute a characteristic feature of this system and explain its higher precision. Thus, measuring performance under conditions of deprived attention may serve as a diagnostic test of whether groupitizing is based on the subitizing system.

Number estimation is not always veridical. The clearest example comes from *numberline* studies, which require participants to map number onto space. Under many conditions, including deprived attention, the mapping shows a strong compressive non-linearity (Cicchini et al., 2014). While this has been described as reflecting a native logarithmic system of encoding number (Cicchini et al., 2014) several recent studies explain the non-linearity as an example of "central tendency" or "regression to the mean", a principle observed in almost all perceptual systems (Hollingworth, 1910). Regression to the mean is well described within the Bayesian framework, where the mean can be considered a Bayesian *prior* (Anobile, Burr, et al., 2019; Cicchini et al., 2014; Jazayeri & Shadlen, 2010). An important prediction from this approach is that the magnitude of the compressive non-linearity should vary with the precision of the numerosity judgments: the worse the precision (higher Weber fractions), the greater should be the non-linearity. If groupitizing is rooted in the subitizing system, which needs attention to boost precision (Burr et al., 2010), we expect there to be less regression to the mean for grouped than ungrouped stimuli, and that this advantage should disappear under attentional deprivation.

In the current study we tested whether the grouping-induced improvements in precision and accuracy of number estimation is based on extending the subitizing system to larger numerosities. To this aim we measured precision and accuracy of numerosity estimation for grouped and ungrouped arrays while modulating attentional resources with dual tasks. If the groupitizing phenomenon is rooted in the subitizing system, attentional deprivation should affect precision more for grouped than ungrouped stimuli. We further explored whether groupitizing may rely on arithmetical computation, with a preliminary study correlating simple calculations skills with precision for estimating grouped or ungrouped numerosities.

4.2 Methods

Sample size was calculated with a Power analyses using G*Power software (version 3.1). As the main goal of the current experiment was to detect a numerosity thresholds change under attentional load the analyses aimed to calculate the required sample size to reliably detect a difference between two dependent means: average Weber Fractions in single and dual task conditions (two tailed paired t-test). The effect size was estimated from Burr et al. (2010). With an $\alpha = 0.05$ and a Power of 0.95, the analyses suggested a required sample size of 6.

Twelve young adults (mean age=26.1, standard deviation=2.9, range= 22-32) participated in this study. Participants were all psychology students with no mathematical or other learning disorders nor over-exercised calculation skills and all with a normal or corrected-to-normal vision.

Stimuli were generated and presented with PsychToolbox(Brainard, 1997) routines for Matlab (ver. R2016b. 9.1.0.441655. The Mathworks, Inc., https://it.mathworks.com). Subjects sat 57 cm from a 19" screen monitor (60 Hz), in a quiet and dimly light room. One experimenter (P.A.M.M.) performed the tests throughout the study. The experimental procedures were approved by the local ethics committee (*Comitato Etico Pediatrico Regionale — Azienda Ospedaliero-Universitaria Meyer*, Florence). The research was performed in accordance with the Declaration of Helsinki and informed consents were obtained from all participants prior to the experiment.

Participants each performed five sessions: in four they were asked to estimate the numerosity of ungrouped or grouped arrays both in single or dual task conditions, while in the fifth session they were given a mental calculation task. The conditions were tested separately with the order counterbalanced across subjects. No feedback was provided, and participants were not informed about the numerosity range. They were also not informed about the different spatial structures of the numerical arrays (ungrouped or grouped), and they were left free to choose any strategy to solve the task, and the possibility of performing mental calculation with the grouped stimuli was never mentioned.

4.2.1 Numerosity Stimuli and Experimental paradigm

Stimuli were the same as those used by Anobile et al. (2020). The arrays were sets of white squares (0.4° × 0.4°) with black borders (in order to balance overall luminance) constrained within a square area of 6° × 6°. The only difference from Anobile et al. (2020) was that in each trial, one item was randomly selected and replaced with a different shape, either a diamond, a triangle or a circle (with a total area equal to that covered by the squares).

In the ungrouped conditions, the position of each item was randomly selected from 106 possible positions within the stimulus area, the centers of equally spread sectors within the 6 × 6 area (each grid 0.5° × 0.5°). For the spatially grouped condition, items were arranged within a maximum of 4 groups (**Figure 4.1**). Each group (spanning over a max area of 1 × 1.5 deg) was located in one quadrant centered at 3° from the central fixation point. Each group was randomly assigned to one quadrant (between 1 and 4), then the individual items positions were randomly selected out the 12 possible locations in the selected quadrant. Within each quadrant, the maximum center-to-center distance between elements was 2° and the minimum was 0.5°.

Each trial started with a black central fixation point that turned white after 1 sec and remained on screen for the entire experiment. After another 1 s an array of items was centrally displayed for 200 ms, followed by a blank screen. In the single tasks (performed separately with ungrouped and grouped stimuli), participants were asked to verbally estimate the numerosity of the array, disregarding the shape of the individual items. The response was entered by the experimenter on the numeric keypad, who also initiated the following trial. Participants were asked to respond quickly, but to concentrate on accuracy. In the dual-tasks (again, performed separately with ungrouped and grouped stimuli) participants were asked first to identify the oddly shaped item by pressing the appropriate arrow key (diamond: left arrow; triangle: down arrow; circle: right arrow), then to verbally estimate the numerosity of the array. The experimenter (blind to the stimuli) hit the spacebar as soon as the response was spelled out, then inserted the number on a numeric pad.

We tested all numerosities between 5 to 17. In the grouped conditions, each numerosity was organized into 2 – 4 clusters, each comprising a variable number of items (between 2 and 6), resulting in the following configurations: 2, 2, 1 – 3, 3 – 3, 3, 1 – 2, 2, 2, 2 – 4, 4 – 3, 3, 3 – 3, 3, 3, 1 – 3, 3, 3, 2 – 3, 3, 3, 3 – 4, 4, 4 – 5, 5, 3 – 4, 4, 3, 3 – 4, 4, 4, 3 – 4, 4, 4, 4 – 5, 5, 6 – 5, 4, 4, 4. All clusters except three (13 = 5, 5, 3; 16 = 5, 5, 6; 17 = 5, 4, 4, 4) contained 1 to 4 elements.

On every trial, numerosities and configuration patterns (i.e. 3,3,3,1 or 3,1,3,3) were randomly selected. Each participant completed 150 trials for each condition, with each numerosity presented in mean 12 times, for a total of 600 trials for the entire experiment. Trials with response times higher than 3 standard deviations were considered outliers and eliminated from the analysis (0.8% of the trails).

4.2.2 Mental calculation test

Mental calculation proficiency was measured by a custom-made computerized test. Each trial started with a central fixation cross. As soon as the participants pressed the space bar, the stimuli (1° ×1.5° digits, and 1° × 1° operand, Arial font) were displayed. Each trial required the participant to mentally solve an arithmetic operation. Each participant solved 37 operations in total. Each operation was randomly selected trial-by-trial between: 3+3, 4+2, 2+5, 3+4, 4+4, 5+3, 3+6, 4+5, 2x3, 2x4, 2x5, 2x6, 2x7, 2x8, 2x9, 3x3, 3x4, 3x5, 3x6, 4x4, 4x5, 4x6, 6-3, 6-4, 7-3, 7-5, 8-3, 8-4, 9-4, 9-6, 2+1+2, 3+1+3, 3+3+3, 3+4+4, 5+3+5, 5+6+5, 6+5+6.

Participants mentally calculated the result as fast as possible and responded verbally (no explicit time limit was provided). The experimenter (blind to the stimuli) hit the spacebar as soon as the participants spelled out the result (which recording response time), then entered the response on the numeric keypad. Trials with response time higher than 3 standard deviations were considered outliers and eliminated from the analysis (1.3% of trails).

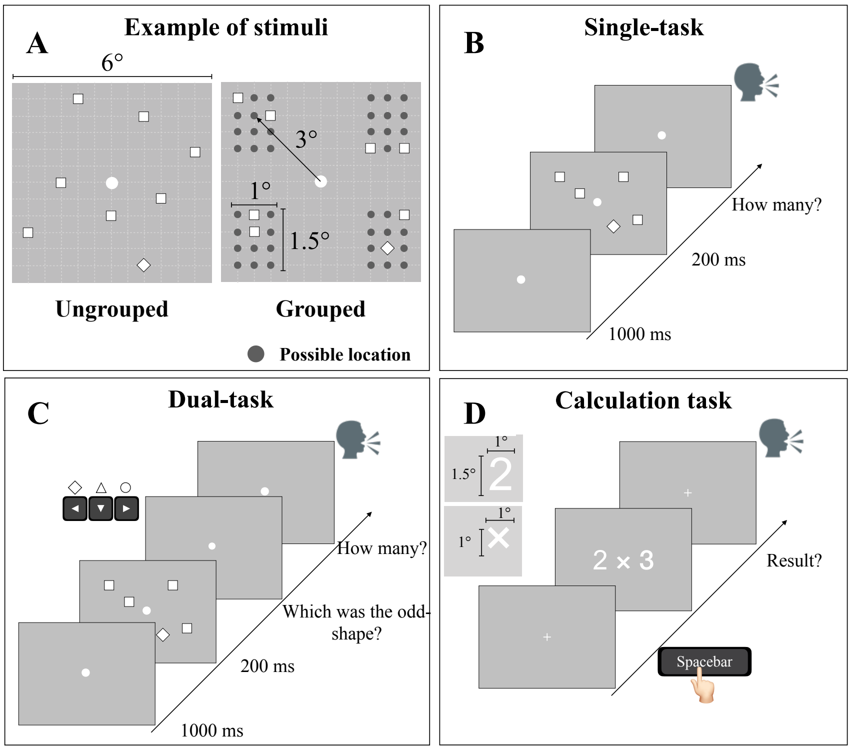


Figure 4.1. Stimuli and procedure. (**A**) Illustration of the procedures followed to generate the stimuli in the ungrouped and grouped conditions. (**B-C**) In the numerosity estimation tasks each trial started with a central fixation point, followed by a briefly flashed ensemble of squared items, with one differing shape (diamond in the example). (**B**) Single-task: Participants were asked to ignore the odd-shaped item and to verbally report the perceived numerosity. (**C**) Dual-task: participants first classified the odd-shape (by appropriate keypress), then verbally reported the perceived numerosity. (**D**) The calculation task started as the participant pressed the spacebar. On every trial, a particular arithmetical operation appeared on the screen (lasting until the response), and participants verbally reported (as fast as possible) the result.

4.2.3 Data analysis

Data were separately analyzed for each subject. For the numerosity estimation task we calculated the average perceived numerosity (accuracy) and the response standard deviation (precision), separately for each numerosity and condition. Standard deviations were divided by the corresponding perceived numerosity, resulting in the Weber fraction (Wf), a dimensionless index of precision (Anobile, Arrighi, et al., 2018). The Weber fractions calculated for each separate numerosity were also averaged across numerosity levels, in order to obtain a summary precision index.

The magnitude of the attentional cost induced by grouped spatial structure was measured as the normalized difference between average Weber fractions calculated in the single (ST) and dual (DT) tasks, averaged across numerosity levels:

$$Attentional\ cost = \frac{Wf_{DT} - Wf_{ST}}{Wf_{DT} + Wf_{ST}} \qquad \text{eqn. (1)}$$

Where Wf_{DT} and Wf_{ST} are average Weber fractions for the dual and single tasks.

The thresholds improvements induced by grouping in the single task was measured as the normalized difference between average Weber fractions calculated in the ungrouped (NG) and grouped (G) conditions, averaged across numerosity levels:

$$Groupitizing\ advantage = \frac{Wf_{NG} - Wf_{G}}{Wf_{NG} + Wf_{G}} \qquad \text{eqn. (2)}$$

Where Wf_{NG} and Wf_{G} are the average Weber fraction for the ungrouped and the grouped conditions in the single task.

Weber fractions were analyzed with Repeated Measures ANOVA and Bonferroni corrected post-hoc t-tests. Effect sizes (η^2 and Cohen's d) are also reported when appropriate. The relation between attentional cost, total numerosity and number of groups was analyzed with zero-order (Spearman) and partial correlations. Log_{10} Bayes factors (LogBF) are reported alongside standard Rho (ρ_s) and p-values. Positive Log_{10} Bayes factors should be interpreted as lending substantial (0.5-1), strong (1-1.5), very strong (1.5-2) and decisive (>2) support to the alternative hypothesis. Negative LogBF within these ranges is evidence for the null hypothesis.

To evaluate non-linear compression of mean estimates of numerosity we fitted the data with power functions:

$$y = aN^b \qquad \text{eqn. (3)}$$

Where y is the average estimate of numerosity, N physical numerosity and a and b constants free to vary. The value of the exponent b is an index of non-linearity, with $b = 1$ implying a linear relationship, and $b < 1$ a compressive non-linearity ($b = 0.5$ implies square root).

The Bayesian central tendency model assumed that the perceived numerosity y was given as a weighted average of the physical numerosity and the mean of the range.

$$y = N\left(1 - w_p\right) + w_p\overline{N} \qquad \text{eqn. (4)}$$

Where w_p is the weight assigned to the prior, which for an optimal observer is proportional to the relative reliabilities (inverse variances) of the two sources of information. Under the simplifying assumption of Weber's Law, this becomes:

$$w_p = \frac{(Wf_i \cdot N)^2}{(Wf_i \cdot N)^2 + \sigma_P^2} \qquad \text{eqn. (5)}$$

Where Wf_i is the Weber fraction for condition, and σ_P^2 is the variance of the prior, estimated to best fit all four conditions simultaneously.

For the mental calculation task, two separate *z* scores were calculated for each participant (using the mean and the standard deviation of the entire group), one for accuracy, the other for response speed. We then averaged the two *z* scores to yield a combined math performance index, following the procedure previously used by Anobile et al. (2018). Participants were categorized as belonging to the "low" or "high" math sample if the combined z-score for mental calculation was below or above the 50th percentile. To evaluate the relation between numerosity estimation and calculation skills we performed standard Pearson' correlations, with correction for multiple comparisons.

Statistical analyses were performed using JASP (version 0.12.2, The JASP Team 2020, https://jasp-stats.org) and Matlab (R2016b).

4.3 Results

4.3.1 Effect of grouping and attention on numerosity estimation thresholds

We used a dual-task paradigm to measure the effect of attentional deprivation on precision and accuracy of numerosity estimation for ungrouped and grouped spatial arrays. Participants estimated numerosity, either during a concurrent visual search task (spot out the odd-shaped item), or with the visual distractor present, but ignored (single-task). **Figure 4.2A** shows that when the distractor was ignored, leaving attentional resources for the numerosity task, there was a strong groupitizing advantage, about 20% on average. Depriving attention affected grouped but not ungrouped stimuli, annulling the groupitizing advantage. For ungrouped stimuli the small effect of attentional deprivation was similar at all numerosities (**Figure 4.2B**), while for grouped stimuli, it was clearly strongest at lower numerosities (**Figure 4.2C**).

These effects were born out by three-way repeated measures ANOVA, with spatial structure (ungrouped or grouped), attentional load (single or double task) and numerosity (13 levels) as factors. There were significant main effects for spatial structure ($F_{(1,11)} = 5.8$, $p = 0.034$, $\eta^2 = 0.013$, $d = 0.23$) and for attentional load ($F_{(1,11)} = 11.2$, $p = 0.006$, $\eta^2 = 0.046$, $d = 0.44$). Crucially, the interaction shown in **Figure 4.2A** between attentional load and spatial structure was significant ($F_{(1,11)} = 5.4$, $p = 0.04$, $\eta^2 = 0.011$, $d = 0.21$). Post-hoc tests showed that with full attention, Weber fractions for grouped arrays were significantly lower than those for ungrouped arrays ($t = 3.35$, $p_{bonf} = 0.017$, squares in **Figure 4.2A**), while in the dual-task they were statistically indistinguishable ($t = 0.11$, $p_{bonf} = 1$). Modulating attention did not alter Weber fractions for ungrouped arrays ($t = 1.37$, $p_{bonf} = 1$) while for grouped arrays, Weber fractions in dual-task were higher than that in single-task ($t = 4.082$, $p_{bonf} = 0.004$). There was also a significant interaction between numerosity and attentional load, being stronger at low numerosities ($F_{(12,132)} = 3.14$, $p < 0.0001$, η^2

= 0.04, d = 0.41). The triple interaction was not significant ($F_{(12,132)} = 0.9$, $p = 0.58$, $\eta^2 = 0.012$, $d = 0.22$). Yet, if groupitizing is based on a capacity-limited, subitizing-like system, depriving attention should most strongly impact the lowest grouped numerosities. Indeed, although the triple interaction did not reach significance, attention seems to affect estimation thresholds more for low numerosities, and only for grouped stimuli. Planned comparison t-tests confirmed that attentional deprivation did not significantly affect estimation thresholds of ungrouped stimuli for any of the numerosities tested (all $p > 0.05$ **Figure 4.2B**). On the other hand, when the stimuli were spatially grouped, attention most strongly modulated estimation thresholds for the lowest numerosity (N5: $t = 5.149$, $p_{bonf} = 0.0007$; N6: $t = 3.913$, $p_{bonf} = 0.158$; N7: $t = 4.48$, $p_{bonf} = 0.015$; $p_{bonf} > 0.05$ for all the other numerosity, **Figure 4.2C,** see also **Figure 4.3B**.

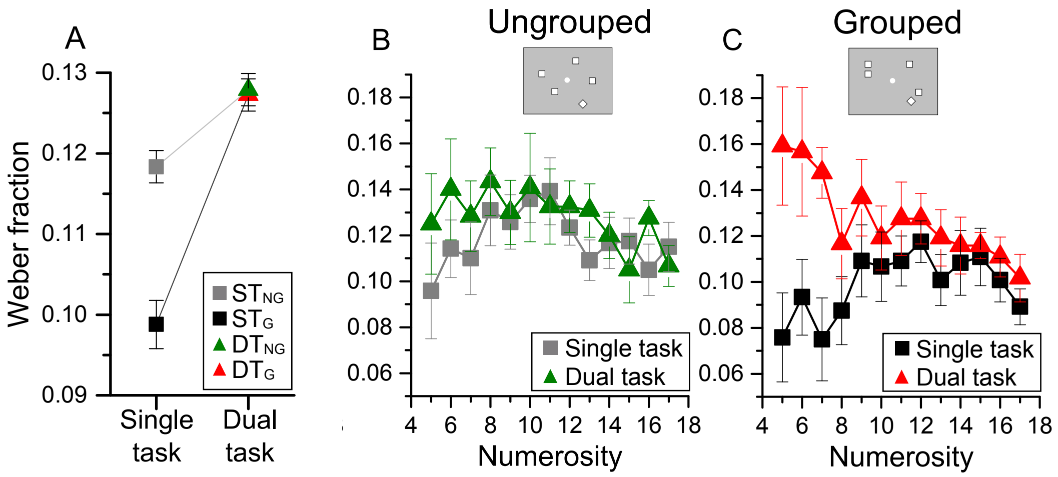


Figure 4.2. Effect of attention on numerosity estimation precision. (**A**) Average Weber fractions for the four conditions showing the interaction between attentional load and stimulus configuration on numerosity estimation. The average Weber fraction for the 4 conditions were: $ST_{NG} = 0.118 \pm 0.002$; $ST_G = 0.099 \pm 0.003$; $DT_{NG} = 0.129 \pm 0.002$; $DT_G = 0.126 \pm 0.002$. (B-C) Average Weber fractions as a function of numerosity plotted separately for ungrouped (**B**) and grouped stimuli (**C**) for the single (squares) and dual (triangles) tasks. Symbols refer to average responses, with error bars ± 1 s.e.m.

To avoid a systematic association between total numerosity and number of groups, numerosities in the grouped condition were presented with different configurations, varying between 2 – 4 clusters. For example, the number eight was shown either with the (2, 2, 2, 2) or with the (4, 4) configurations. We tested whether the attentional modulation of thresholds was particularly marked for certain configurations, and whether it depended primarily on the number of groups or on the total numerosity, or both. We correlated the attentional cost (defined as the normalized difference between Weber fractions in the single and dual conditions: eqn.1) with the number of groups and total numerosity (**Figure 4.3**). As larger numerosities were generally divided into more groups than lower numerosities (positive correlation between total numerosity and number of subgroups: $\rho_s = 0.51$, $p = 0.02$, LogBF = 0.8), we also calculated partial correlations, evaluating the variance independently explained by each of these factors (total numerosity or number of groups). Attentional cost negatively correlated with both the number of groups and total numerosity (both $\rho_s < 0.001$, LogBF > 1.7), suggesting that the detrimental effect of attention was higher when both the number of groups and the total numerosity were lower and tended to decrease for larger numerosities. The correlation between the attentional cost and total numerosity remained significant even when taking into account the effect of the number of groups ($\rho_s = -0.53$, $p = 0.017$, LogBF = 0.90). Similarly, the correlation between attentional cost and number of groups also remained significant when controlling for the total numerosity ($\rho_s = -0.62$, $p = 0.006$, LogBF = 0.99). These results indicate that attentional deprivation acts on both the total numerosity and on the number of groups: its negative impact on estimation thresholds was strongest for the lowest numerosities and for stimuli divided into fewer groups.

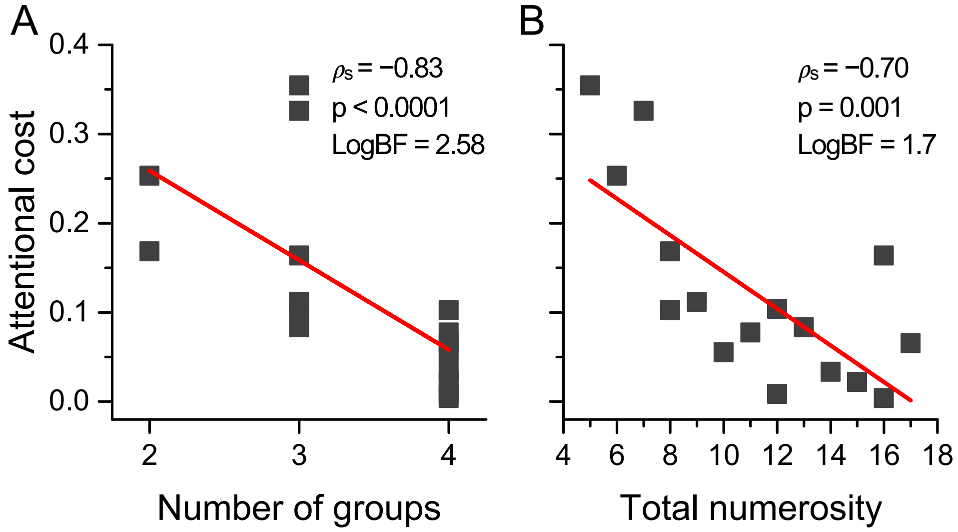


Figure 4.3. Relationship between attentional cost, number of groups and total numerosity. Attentional cost correlated with the number of groups (**A**) and with the total numerosity (**B**).

4.3.2 Effect of spatial structure and attention on perceived numerosity

Under many conditions, including deprived attention, the mapping shows a strong compressive non-linearity (Cicchini et al., 2014), considered by many as an example of regression to the mean. If groupitizing is rooted in the attention-dependent subitizing system, which requires attention to boost numerical estimation precision, the effects of grouping and attentional deprivation should also be evident in estimation accuracy.

Figure 4.4 (A-D) shows the average estimates of numerosity for the four conditions. In general, low numerosities were overestimated and high numerosities underestimated, both following a regression to the mean. However, as usually observed, the regression to the mean was greater at high numerosities (where precision is less), resulting in a strong compressive non-linearity. To measure the non-linearity created by these biases, we fitted each set of data with a power function (eqn. 3, methods), shown by the blue lines. The fits were all very good (total R^2 over all conditions = 0.986).

Importantly, as predicted, the non-linearity was not the same in all four conditions, but was highest for conditions with the highest Weber fractions. **Figure**

4.4E plots the index of the power function against average Weber fraction. The non-linearity clearly increases with Weber fractions, from 0.89 for the grouped single task condition (index of 1 means a linear function), to 0.80 for the ungrouped single task condition to 0.72 for the two dual task conditions. Where performance is most precise, it is also most accurate. The correlation between the two measures was $r = -0.983$, $p = 0.008$, LogBF = 0.84.

To test the quantitative predictive power of the Bayesian model of central tendency, we fitted the data with the Bayesian prediction, given by eqn. 4 of methods. The equation essentially states that perceived numerosity will be a weighted average of the actual physical numerosity of the stimulus and the mean numerosity of the range tested (the *prior*). Relative weighting of the two is determined by their precision: the more precise the estimates, the higher the weighting (eqn. 5). That has two consequences. Assuming constant Weber fractions implies that thresholds increase linearly with numerosity, so the regression effects will be more pronounced at higher than at lower numerosities, leading to the compressive non-linearity. Secondly, as the Weber fractions increase between conditions, the prior (which we assume to remain constant between conditions) will have greater effect, resulting in the greater non-linearities that we observe (**Figure 4.4E**).

The fits are shown by the red curves of **Figure 4.4A-D**. The four fits have only 1 degree of freedom for all of them, the width of the prior (σ_P of eqn. 5) was constant for all four conditions, selected to simultaneously minimize the residuals of all four fits. The resulting fits were excellent, with total $R^2 = 0.988$ (compared with 0.986 for the power fits). Thus, the Bayesian central tendency model explains well the data, qualitatively and quantitatively.

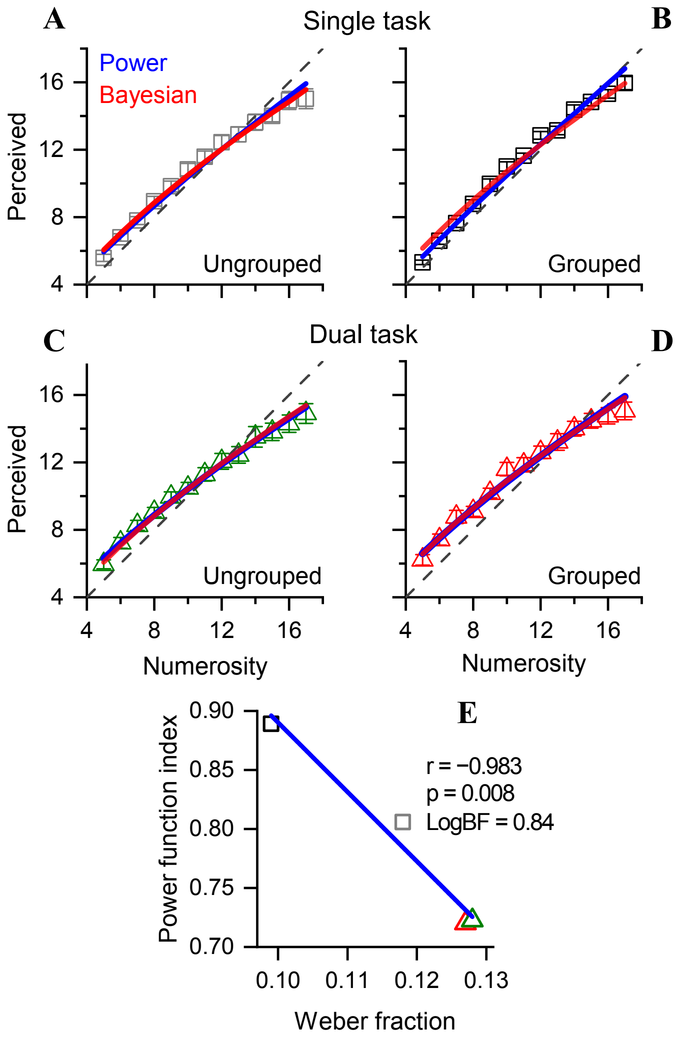


Figure 4.4. Perceived numerosity. Perceived numerosity as a function of physical numerosity for estimation of ungrouped (**A**) or grouped (**B**) stimuli in single and dual task (**C-D**). Continuous lines are the best fit of power function (blue) and Bayesian modelling (red). (**E**) Power function index correlated with the average Weber fraction. Symbols refer to average across participants, with error bars ± 1 s.e.m.

4.3.3 Relation with arithmetical abilities

Despite the relatively small number of participants in this study (primarily designed to examine in detail the effects of attention on groupitizing), we also looked for possible correlations between groupitizing and math skills. Participants did a simple speeded calculation test described in methods, which was scored for both speed and accuracy. The average accuracy across participants 90%±7%, and average speed 1.3±0.3 secs. We combined z-scores of speed and accuracy (see methods) and correlated this index against Weber fractions for ungrouped and grouped stimuli. For ungrouped stimuli, Weber fractions were uncorrelated with the math index ($r = -0.18$, $p = 0.288$, LogBF = −0.24; **Figure 4.5A**); but for grouped stimuli the correlation was significant, and remained close to significance after correcting for multiple comparison ($\alpha = 0.5/2$: $r = -0.56$, $p = 0.029$, LogBF = 0.54; **Figure 4.5B**). We also hypothesized that participants with higher arithmetical skills would gain more from grouping of stimuli than less skilled participants, which was verified by the positive correlation between “grouping advantage” and math index ($r = 0.58$, $p = 0.023$, LogBF = 0.61; **Figure 4.5C**). While these results should be taken with caution before replication in future studies, they suggest the very interesting possibility that groupitizing could be a sensitive predictor of math skills.

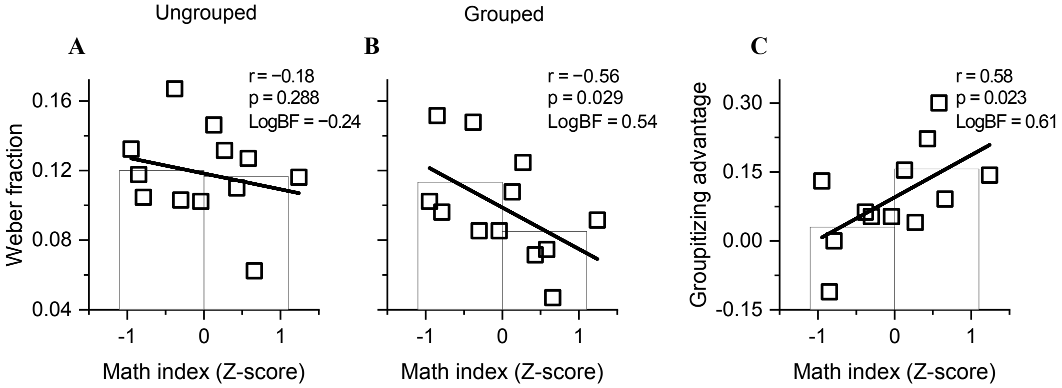


Figure 4.5. Relation between estimation precision and mental calculation abilities. (**A-B**) Weber Fractions plotted against math index for all participants. Bars show averages for median split. The correlation was insignificant for ungrouped but significant for grouped stimuli (values in graph and main text). (**C**) Groupitizing advantage as a function of math index. The correlation was positive and significant (values in graph and main text).

4.4 Discussion

The aim of the present study was to directly test whether the *groupitizing* phenomenon (G. S. Starkey & McCandliss, 2014) depends on subitizing, by measuring the consequences of depriving attentional resources on numerosity estimation thresholds of spatially grouped and ungrouped items. As previous studies (Anobile, Castaldi, et al., 2020) have shown, numerosity thresholds for spatially grouped stimuli were lower than for randomly scattered stimuli. However, depriving attention with a concomitant dual task completely obliterated the groupitizing advantage, consistent with the suggestion that it relies on subitizing. We also explored the link between groupitizing and arithmetic, and showed that simple mental calculations skills in adult participants correlated with estimation thresholds for grouped but not ungrouped stimuli, and also with the advantage given by grouping.

Although subitizing was originally thought to be pre-attentive, dependence on attention has become a signature of the subitizing system. Many studies have shown that attention has a much stronger detrimental effect in the subitizing than estimation range, enough to equate subitizing precision and reaction times to those of higher numerosities during dual tasks (Anobile et al., 2012; Burr et al., 2010; Pomè, Anobile, Cicchini, Scabia, et al., 2019; Railo et al., 2008; Vetter et al., 2008). The selective detrimental effect of attentional deprivation in the subitizing range was reinforced by a recent clinical single case study with a simultanagnosic patient (Anobile, Tomaiuolo, et al., 2020), who suffered a severe visual attentional deficit. PA showed no subitizing advantage for low numerosities, while his numerosity perception was relatively spared for intermediate numerosities, above the subitizing range. The subitizing advantage, at least in the visual domain, could thus emerge from the well-known capacity-limited attentive tracking system, that allows precise tagging of a few objects in space (Piazza, 2010). Other studies show that depriving auditory and haptic attentional resources also affects visual subitizing (Anobile et al., 2012). Future studies should investigate the effect of cross-modal attention deprivation on groupitizing.

The current study showed that performing a dual task completely eliminates the groupitizing advantage for estimation thresholds, in the same way that it eliminates the subitizing advantage for low numbers: estimation thresholds for grouped arrays in dual task became like those measured with ungrouped arrays in single task. Depriving attention during estimation of ungrouped arrays, on the other hand, did not affect estimation thresholds. Given that the numerosities tested were the same across the grouped and ungrouped conditions (in both cases well exceeding the subitizing range), the only factor driving the attentional modulation was the spatial configuration. We presume that ungrouped arrays were judged primarily by estimation system, largely independently of attention, whereas grouped arrays trigger the additional intervention of the subitizing system, which boosts performance. However, as subitizing requires attentional resources, during dual-task only the estimation system could operate, bringing performance for grouped arrays down to that of ungrouped stimuli. In the grouped condition, the detrimental effect of dual task scaled both with total numerosity and with the number of groups, with stronger cost for low numerosities and lower number of groups. The higher cost of attention for low numerosities and fewer groups suggests that groupitizing acts on both these factors. With larger total numerosities and/or number of groups, the attentional free estimation system is likely to kick in, even if items are spatially segregated, resulting in a weaker attentional modulation of estimation thresholds.

We also found that estimation biases differed across attentional and grouping conditions. All estimates departed from linearity and tended toward the center of the numerosity range, with the effect increasing when attention was deprived. The observed compressed non-linearity was well fitted by a Bayesian model of central tendency (Alais & Burr, 2004; Cicchini et al., 2012; Hollingworth, 1910; Jazayeri & Shadlen, 2010). This effect has been described for a wide range of stimuli (Alexi et al., 2018; Cicchini & Burr, 2018; Liberman et al., 2014; St. John-Saaltink et al., 2016; Taubert et al., 2016; Xia et al., 2016), and is thought to maximize the perceptual efficiency by exploiting contextual effects. An important prediction of the Bayesian model is that the magnitude of the non-linearity should depend on perceptual thresholds. This prediction was borne out, with a strong and significant correlation between magnitude of non-linearity and Weber fractions. And the Weber fractions predicted well the form of the non-linearity, with only one degree of freedom (strength of the prior, unchanged between conditions).

We further explored whether groupitizing may depend on the ability to make simple calculations on grouped stimuli (Anobile, Castaldi, et al., 2020;

Ciccione & Dehaene, 2020; G. S. Starkey & McCandliss, 2014). The correlation between arithmetic skills and Weber fractions of grouped (but not ungrouped) stimuli, and also with the groupitizing advantage in our small sample suggests that this may be the case. We emphasize, however, that although estimation thresholds of ungrouped arrays were uncorrelated with math ability in our small sample of adults, we do not believe that this contradicts theories suggesting that an efficient Approximate Number System (ANS) may be a pre-requisite for typical development of math skills (Dehaene, 1992; Halberda et al., 2008; Piazza, 2010). The link between ANS and math abilities is much less evident in adults than in children (Anobile, Arrighi, et al., 2018; Castaldi et al., 2018; Castaldi, Turi, et al., 2020; Inglis et al., 2011). Many studies have reported that numerosity perception precision sharply improves during development and formal arithmetical learning (Halberda et al., 2012; Halberda & Feigenson, 2008; Libertus et al., 2012; Odic et al., 2013; Piazza et al., 2013; but see also: Holloway & Ansari, 2009; Iuculano et al., 2008; Sasanguie et al., 2013), while in educated adults, symbolic math abilities may be already steadily mapped into their basic non-symbolic representation, making the association less evident (Anobile, Arrighi, et al., 2018; Braham & Libertus, 2018; Feigenson et al., 2013; Inglis et al., 2011; Krueger, 1984). While ANS precision measured with ungrouped stimuli may be a reliable predictor of early math abilities in childhood, once the number acuity has refined and been mapped onto symbolic numbers, it could lose part of its predictive power. However, groupitizing relies less on approximate numerical estimation, but triggers calculation strategies to combine subitized subsets. This was confirmed by participant subjective reports. Although grouping strategies were never mentioned in participant instructions, when debriefed all participants reported to have used arithmetical strategies (addition and in some cases multiplication of the subgroups). Participants also reported that they had more difficulties in applying these strategies when the stimuli were ungrouped. In this condition, participants may have used a combination of different approaches, probably weakening the link with mental calculation skills.

Importantly, the efficiency of the subitizing system by itself-seems may not to be sufficient to predict calculation skills. Previous studies found no significant correlation between subitizing capacity and math skills in children or adults (Anobile, Arrighi, et al., 2019). Moreover, while the subitizing system is already functional as early as 2 years of age (Klein & Starkey, 1988), 6-year-old preschoolers cannot take advantage of groupitizing (G. S. Starkey & McCandliss, 2014). Thus, the relationship between groupitizing and arithmetic is most likely driven by using calculation skills to extend the subitizing range, rather than on the capacity to subitize. It should be mentioned, however, that exact serial counting speed has been shown to be a good marker of arithmetical abilities (Gray & Reeve, 2014; Reeve et al., 2012), leaving open the possibility that the link between arithmetic and subitizing may emerge more clearly when slow counting is used instead of fast approximation, as in the current study. Also, a recent study on kindergarten children has suggested that subitizing may play a role in the development of symbolic number abilities, opening the possibility that the link would be stronger in the earliest developmental stages (Hutchison et al., 2020).

In this study, like previous studies, we deliberately facilitated the use of grouping strategies by spatially grouping the stimuli. Other manipulations also aid grouping, such as organizing stimuli into same-coloured groups. It would be interesting to explore further what other organizations may encourage groupitizing. For example, mirror symmetry biases numerosity estimates, so symmetrical patterns appear less numerous than their asymmetric counterparts (Apthorp & Bell, 2015). It is possible that symmetry would also facilitate grouping, leading to lower thresholds. This would be well worth exploring, together with other manipulations of shape and organization.

While the correlational results of this study should be taken with some caution, given the small number of participants, our explorative analysis should encourage future work investigating whether numerosity thresholds measured with grouped arrays (using a variety of grouping cues) may prove to be a more sensitive predictor of arithmetical abilities in adults. These studies should also explore the contribution of other domain general processes, such as attentional and working memory resources to the groupitizing advantage and their predictive role with different components of the arithmetical competence.

Chapter 5

Numerosity estimation of grouped and ungrouped arrays: an fMRI study

Chapter 5

5.1 Introduction

In the first part of the thesis, I described the studies that explored the neural substrate underlying the ‘number sense’. These studies suggested that non-symbolic numerosities are processed by a dedicated brain circuitry that engages areas in a fronto-parietal network (Ansari & Dhital, 2006; Dormal & Pesenti, 2009). This network is sometimes described as being more right-lateralized (Dormal et al., 2010; Pinel & Dehaene, 2010), especially in children (Cantlon et al., 2006; Izard et al., 2008) and gradually involves more bilateral regions with increasing practice with Arabic digits and arithmetical skills (Cantlon et al., 2006). Indeed, arithmetical calculation was found to heavily recruit the fronto-parietal network in the left hemisphere (Pinel & Dehaene, 2010).

Recently, one study described the neural substrate of addition testing both symbolic and non-symbolic formats (Bugden et al., 2019). They found that both formats engaged the bilateral IPS relative to color control tasks. However, symbolic addition elicited a left lateralized network including left precentral gyrus while non-symbolic addition activated small clusters in the occipital lobe. In addition common neural activations for non-symbolic and symbolic addition were found in the superior parietal lobule and in the bilateral inferior temporal gyri (Bugden et al., 2019).

Neuropsychological cases of double dissociations between the ability to solve multiplications and subtractions (reviewed in: Dehaene et al., 2003) have led to the suggestion that multiplications may be typically solved by recalling the solution from rote verbal memory, whereas subtractions may require actual computation based on some sort of internal manipulation of numerical quantities on an internal number line, possibly similar to the strategy employed to solve numerical comparisons. When tasks put greater requirement on verbal encoding of numbers and retrieve arithmetic facts, the left angular gyrus shows increasingly greater activation. For example, for exact calculation, the left angular gyrus, as well as inferior temporal gyrus shows greater activation for operations that require access to a rote verbal memory of arithmetic facts, than for operations that are not memorized and require quantity manipulation.

In the previous chapters, we found that groupitizing depends on subitizing and on calculation abilities (Maldonado Moscoso et al., 2020). It was suggested that groupitizing strategies required participants to subitize the items in each subgroup and then combine the group estimates through multiplication-and-sum strategies. Here, we explicitly tested the hypothesis that having to estimate the numerosity of grouped rather than ungrouped stimuli may automatically trigger arithmetical calculation and therefore recruit the network typically involved during calculation tasks compared to the one typically activated for non-symbolic numerosity perception and estimation.

5.2 Methods

5.2.1 Subjects and MRI acquisition procedure

Fifteen adult volunteers (10 males and 5 females, 28,2 ± 6,2 years old) with normal or corrected to normal vision participated in the study. This study was approved by the ethical committee of the University of Regensburg and all participants gave written informed consent prior to the study. Due to technical problems during the data acquisition, one subject could not complete the study and was therefore discarded from the analysis.

Functional images were acquired on a Prisma 3T scanner (Siemens, Erlangen, Germany) using 64 channel head coil (Erlangen, Germany) as T2*-weighted fast-saturation echo-planar image (EPI) volumes with 2mm isotropic voxels (repetition time [TR] = 2000 ms, echo time [TE] = 30, flip angle [FA] = 52).

T1-weighted anatomical images were acquired at 0.8 mm isotropic resolution (repetition time [TR] = 2400, echo time [TE] = 2.18, flip angle [FA] = 8; 208 transversal slices were acquired). During the scanning head movements were minimized by padding and tape. Visual stimuli were viewed through a mirror anchored to the head coil, which back-projected the stimuli displayed onto a translucent screen located at the end of the scanner bore. Participants were asked to provide occasional responses by pressing one of three buttons on an MRI-compatible response box.

5.2.2 Stimuli and experimental design

During the fMRI scanning participants were centrally presented with arrays of white items (0.4° x 0.4°) with black borders displayed on a grey background, with the overall luminance balanced. The stimuli were created as in Maldonado Moscoso et al. (2020): all but one items were squares, and one odd shape was randomly selected to be either a diamond, a triangle or a circle (with total area matched to that of the squares). Each array comprised 8, 12 or 16 items (Figure 5.1B). Item location within each array was either randomly selected from 106 possible coordinates within the 6° x 6° stimulus area (ungrouped spatial arrangement) or selected within a maximum of 4 groups (each group spanning 1° x 1.5° and being located at 3° from the central fixation point), so that the items were grouped (grouped spatial arrangement). In the latter case, locations of individual items were selected out of the 12 possible coordinates included in the selected quadrant and each group of items was randomly assigned to one quadrant. Grouped stimuli were created with two different configurations, so that there was no systematic association between the numerosity shown within each group and the overall numerosity.

Each run was divided into two parts in which participants performed a numerosity and a shape control estimation tasks on exactly the same stimuli, following the instructions (Figure 5.1A). The two tasks were performed either in the first or in the second half of the run with counterbalanced order. Instructions were shown for 2 s and specified whether participants had to attend to the number of items (number task) or to the odd-shape presented on every trial (shape control task). Four seconds after the instructions, the first array was briefly presented for 200 ms and participants were instructed to attend to the cued dimension and to hold this information in memory until the following trial was presented. After a variable ISI of 3.8 - 5.8 s, either a new array or a question mark was presented. If a new array was presented, participants had to update their memory with the new stimulus and no response was required. If a question mark appeared, participants had to either estimate the numerosity or the shape of the odd item, depending on the task. Responses were provided by pressing one out of three buttons of the response box.

For each task, both grouped and ungrouped arrays were shown in different blocks, however the spatial arrangement change was not explicitly signaled (Figure

5.1A). For each task 48 trials were presented: 6 trials for each of the 6 conditions (3 numerosity x 2 spatial arrangements) and 12 question marks.

Within each scanning session participants performed eight runs of ~10 min. Each run included 8 blocks where the two tasks alternated. The type of task the run started with was balanced across runs and participants.

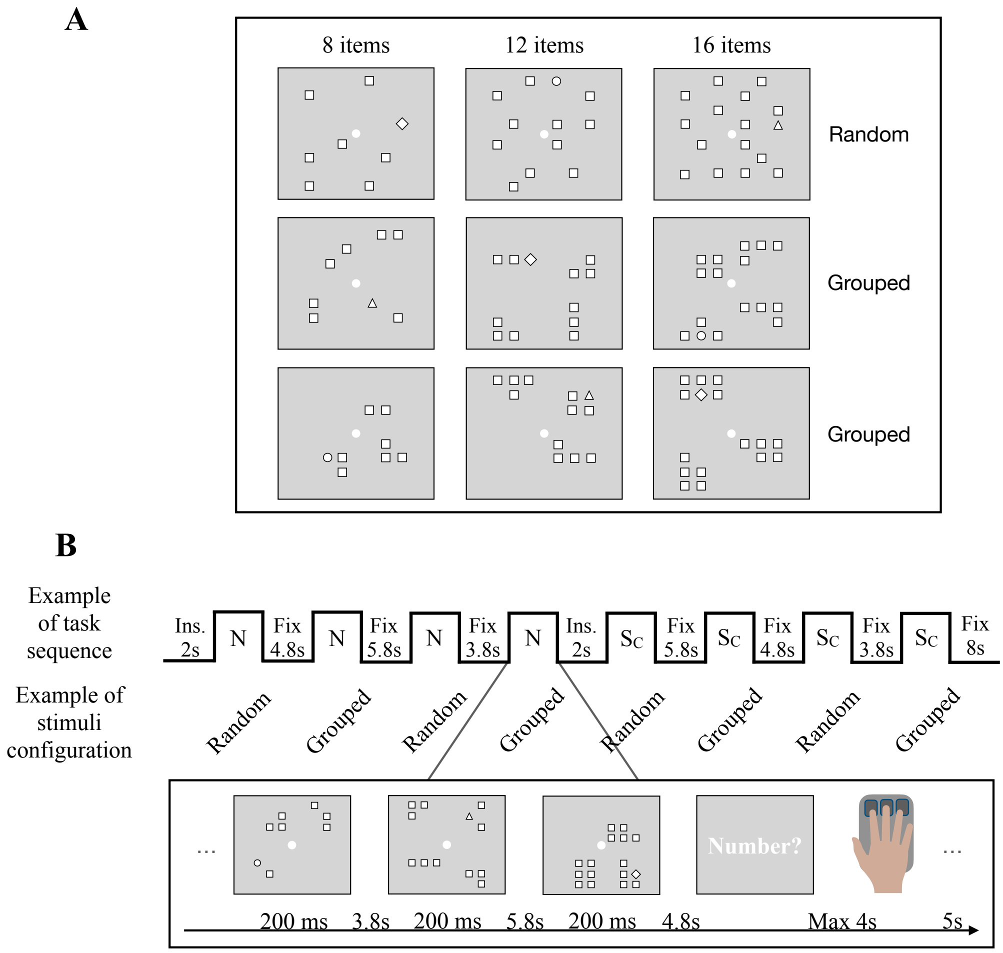


Figure 5.1. Overview of the experimental design and example of stimuli configuration. (A) Example of stimuli configurations. (B) One run comprised 8 blocks in which stimuli configuration could be ungrouped or grouped.

Prior to the scanning participants were tested with the behavioral numerosity estimation task of Maldonado Moscoso (2020). Participants sat at 57 cm from a

19’’ screen monitor (60 Hz), in a quiet and dimly light room. Participants were asked to estimate the numerosity of grouped (3 blocks) or ungrouped (3 blocks) arrays. The presentation of the conditions was counterbalanced across participants and again participants were not informed about the different spatial arrangement of the stimuli. The duration as well as the dimensions of the arrays were the same used during MRI acquisition.

Each trial started with a black central fixation point that turned white after 1s and remained on screen for the entire experiment. The first array was centrally presented after 1 s and was followed by a blank screen. Participants were asked to verbally estimate the numerosity of the array, neglecting the shape of the individual items, but concentrating on providing the numerical estimation as quickly and accurately as possible. The experimenter entered the response on the numerical keypad and initiated the following trial. Numerosities between 5 to 17 were tested. In the grouped conditions, each numerosity was organized into clusters (between 2 and 4), each containing a variable number items (between 2 and 6), resulting in the following configurations: 2, 2, 1 – 3, 3 – 3, 3, 1 – 2, 2, 2, 2 – 4, 4 – 3, 3, 3 – 3, 3, 3, 1 – 3, 3, 3, 2 – 3, 3, 3, 3 – 4, 4, 4 – 5, 5, 3 – 4, 4, 3, 3 – 4, 4, 4, 3 – 4, 4, 4, 4 – 5, 5, 6 – 5, 4, 4, 4. All clusters except three (13 = 5, 5, 3; 16 = 5, 5, 6; 17 = 5, 4, 4, 4) contained from 1 to 4 elements. The numerosities as well as the configuration patterns were randomly selected on every trial. Each participant performed 150 trials for each condition (ungrouped and grouped), with each numerosity presented on average 12 times, for a total of 300 trials for the entire numerosity estimation task.

Stimuli were generated and presented using PsychToolbox routines (Brainard, 1997), operating under Matlab (ver. R2016b. 9.1.0.441655. The Mathworks, Inc., https ://it.mathworks.com).

5.2.3 Data analysis

EPI images were preprocessed and analyzed with FSFAST tools of Freesurfer 6.0 (https://surfer.nmr.mgh.harvard.edu/). Preprocessing included motion correction and smoothing with a 3D Gaussian kernel (FWHM = 5 mm). We performed surface-based reconstruction and individual participants' data were sampled to the left and right hemisphere of fsaverage (the surface area of the Freesurfer average subject).

The preprocessed EPI images were entered into two general linear models estimated on subject's surface space. Predictors were convolved with the SPM canonical hemodynamic response function.

In the first General Linear Model (GLM), for each participant we separately modelled the effects of the 4 conditions (2 spatial arrangements x 2 tasks), the instruction and the response trials. To identify which brain regions were involved in numerosity estimation of ungrouped and grouped arrays, for each participant we contrasted the activity elicited during the number task (separately when performed on ungrouped and grouped arrays) against that elicited during the respective control conditions (i.e., 'Ungrouped number task > Ungrouped shape control task'; and 'Grouped number task > Grouped shape control task'). We then performed a random effects group analysis. The resulting statistical maps were thresholded at $p < 0.001$, using correction for multiple comparison at cluster level (Hagler et al., 2006) with cluster forming threshold $p < 0.001$. Next, to reveal regions that responded more to the number task both when the arrays were ungrouped or grouped than to their respective control conditions, we performed a conjunction of random effect analysis across the two contrasts, i.e., '(Grouped number task > Grouped shape control task) ∩ (Ungrouped number task > Ungrouped shape control task)'. This analysis showed the brain regions that were activated for both contrasts (not just one or the other). Finally, for each participant we determined whether there were brain regions that showed greater activation specifically (and not shared) for the number task when the arrays were grouped or ungrouped after subtracting out activity associated with their control tasks, i.e. '(Grouped number task > Grouped shape control task) > (Ungrouped number task > Ungrouped shape control task)'; and '(Ungrouped number task > Ungrouped shape control task) > (Grouped number task > Grouped shape control task)'. We then performed a random effects group analysis, as described above.

In the second GLM we additionally modelled the effect for each numerosity separately, resulting in 14 predictors: 12 conditions (3 numerosities x 2 spatial arrangements x 2 tasks), the instruction and the response trials. The beta estimates for the 12 conditions were entered into pattern recognition analysis. For each participant we anatomically defined regions of interest (ROIs) from V1 to IPS5 derived from a surface based probabilistic atlas (L. Wang et al., 2015) and other ROIs derived from the Freesurfer atlas (Destrieux et al., 2010). ROIs were created on Freesurfer surface and were back projected onto each participant's volume space. Left and right hemisphere for each ROI were merged. ROIs from V1 to IPS5 were further merged in three ROIs corresponding to early (V1 to V3), intermediate (V3A, V3B and V7 also known as IPS0) and higher-level (IPS1 to IPS5) ROIs. Then, we also defined a region called IPS excluding IPS0-5 which was defined by excluding the ROIs from IPS0 to IPS5 from the intraparietal and transverse parietal sulci ROI as defined by the Freesurfer atlas. This region was found to be specifically involved during calculation and numerosity comparison as opposed to numerosity perception (Castaldi, Vignaud, et al., 2020). The parietal and angular gyrus (from now on referred as angular gyrus for brevity) ROI was defined based on the Freesurfer atlas.

Within each of these bilateral ROIs we selected on a subject-by-subject basis an equal number of 1000 voxels that responded most strongly to the orthogonal contrast ('all numerosities > baseline') for pattern recognition analysis. Pattern classification analysis was performed in sckit-learn (Pedregosa et al., 2011) using beta estimates after subtracting the voxel-wise mean across conditions. Linear support vector machines (SVM) with regularization parameter C = 1 was applied. Classification analysis was performed following a leave one run out cross-validation scheme. Classification accuracy obtained from each cross-validation cycle were then averaged together. Pairwise classification was performed for all pairs of numerosities, keeping patterns separated by task (always selecting trials recorded during the number task only) and by spatial arrangements (Train ungrouped -Test ungrouped; and Train grouped-Test grouped). We then also tested for the ability of the classifier to generalize across spatial arrangements. Classification accuracy was then averaged across all pairs of numerosity. Significance against the theoretical chance level (50%) was tested with one-sample t-test and significance was reported after correction for multiple comparisons, as well as log10 Bayes Factor (logBF). Bayes factors should be interpreted as lending positive (< 0.5), substantial (0.5 - 1), strong (1 - 1.5), very strong (1.5 - 2) support to the alternative hypothesis. Repeated measure ANOVA and Bonferroni corrected t-test were performed on classification accuracy with ROIs and spatial arrangements as factors. Greenhouse-Geisser was applied when sphericity was violated. Effect size (η^2) was also reported when appropriate.

From psychophysical data, we calculated the average perceived numerosity and the response standard deviation, separately for each participant, numerosity and condition. We then calculated Weber fraction (Wf) by dividing the standard deviations by the corresponding perceived numerosity. The Wfs calculated for each numerosity were then averaged across numerosity levels in order to obtain two summary precision indices for each participant: one for ungrouped condition and the other one for grouped condition.

Wfs were analyzed with a paired samples t-test. Effect size (Cohen's d) were also reported. We also performed Pearson's correlations between the behavioral Weber fraction and classification accuracy in the V1-V3 and angular gyrus ROIs.

5.3 Results

5.3.1 Psychophysical performance

Before scanning, participants were tested with a behavioral experiment to measure Weber fractions for numerosity estimation of stimuli with different spatial arrangement. We replicated the single task used in Maldonado Moscoso et al. (2020) by asking participants to estimate the numerosity of ungrouped and grouped arrays. Wfs for ungrouped and grouped stimuli were compared with paired sample t-tests. The results support previous evidence (Maldonado Moscoso et al., 2020) showing that Wf were lower when stimuli were grouped compared to when they were ungrouped ($t = -7.52$, $p < 0.0001$, Cohen's $d = -2.08$).

5.3.2 Behavioral performance during scanning

Behavioral performance measured during the scanning was analyzed by repeated measures ANOVA with task (numerosity and shape control tasks) and spatial arrangement (ungrouped and grouped) as within subject variables. ANOVA were performed separately on accuracies and RTs as dependent variables. The interaction of task and spatial arrangements was significant for accuracy ($F(1,13) = 13.257$, $p = 0.003$, $\eta^2 = 0.141$). Post-hoc comparisons revealed that when estimating ungrouped arrays, the accuracy in the numerosity task was lower than in the shape control task ($t = 4.778$, $p = 0.0004$), suggesting that the latter task was slightly easier, although accuracy was very high in both cases (accuracy ungrouped numerosity task: 75%; accuracy ungrouped shape control task: 86%; Figure 5.2A). On the contrary when estimating grouped arrays, the accuracy between the numerosity and shape control task was not statistically different ($t = 0.37$, $p = 0.99$) and was in both cases very high (accuracy grouped numerosity task: 78%; accuracy grouped shape control task: 79%; Figure 5.2A). The accuracy for performing the numerosity task on ungrouped and grouped arrays as well as the shape control task on ungrouped and grouped arrays were not statistically different (ungrouped and grouped numerosity task: $t = 1.473$, $p = 0.92$; ungrouped and grouped shape control task: $t = 2.739$, $p = 0.069$).

Neither the main effects nor the interaction between task and spatial arrangement (Figure 5.2B) were statistically significant for RTs (interaction: $F(1,13) = 2.522$, $p = 0.136$, $\eta^2 = 0.033$, main effect of task: $F(1,13) = 1.09$, $p = 0.315$, $\eta^2 = 0.05$; main effect of spatial arrangement: $F(1,13) = 0.895$, $p = 0.361$, $\eta^2 = 0.01$).

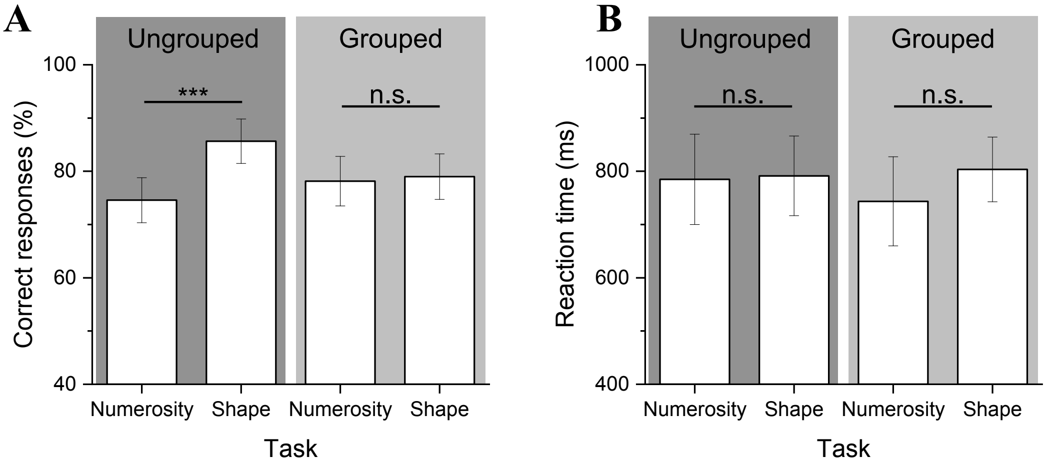


Figure 5.2. Behavioral performance during scanning. (A) Percentage of correct responses and average reaction time (B) to match stimuli for the two tasks in both ungrouped and grouped spatial configuration. Error bar represent ±1 s.e.m. ***p < 0.001.

5.3.3 Univariate analysis

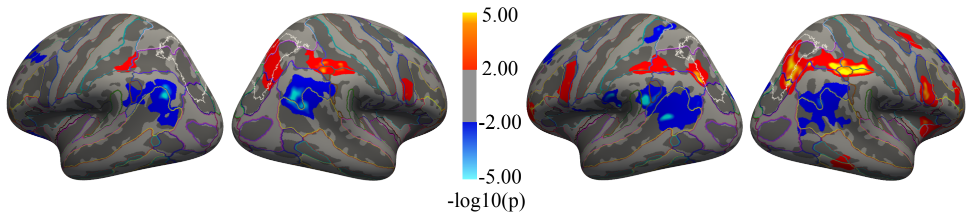
The contrast 'Ungrouped numerosity > Ungrouped shape control' (red activations in figure 5.3A) revealed greater activation for processing numerosity compared to the shape control task of ungrouped stimuli in parietal and frontal cortex. Specifically, activation covered the right superior and transverse occipital sulcus, right intraparietal and transverse parietal sulci (IPS), right superior parietal gyrus (SPG) and in the bilateral postcentral sulcus. In the frontal cortex activation covered the right inferior precentral sulcus and superior frontal gyrus.

The contrast 'Grouped numerosity > Grouped shape control' (red activations in figure 5.3B) elicited activation in similar regions, however additionally recruiting the left hemisphere much more. Additional activations were also observed in right insula and in right inferior temporal gyrus.

The contrast, 'Ungrouped shape control > Ungrouped numerosity' and 'Grouped shape control > Grouped numerosity' (blue activations in figure 5.3A&B)

elicited activation in several bilateral parieto-occipital and superior temporal areas. Widespread activations were also observed bilaterally in the frontal cortex (mainly in the superior and middle frontal gyrus and cingulate gyrus).

The conjunction analysis '(grouped number task > grouped shape control task) ∩ (ungrouped number task > ungrouped shape control task)' highlighted the regions commonly activated when performing the number task on both grouped and ungrouped stimuli relative to their respective control tasks. The results shown in figure 5.2C revealed that common neural activation was found for grouped and ungrouped number tasks in the right superior occipital sulcus and transverse occipital sulcus, right intraparietal sulcus and transverse parietal sulci (IPS), right superior parietal gyrus (SPG) and in the postcentral sulcus bilaterally. Activations were observed also in the frontal cortex, specifically in the right precentral sulcus and right superior frontal gyrus. The results suggest that these regions play an important role in the estimation of non-symbolic quantities irrespective of spatial configuration.



A Ungrouped numerosity > Ungrouped shape control

B Grouped numerosity > Grouped shape control

C Grouped numerosity > Grouped shape control ∩ Ungrouped numerosity > Ungrouped shape control

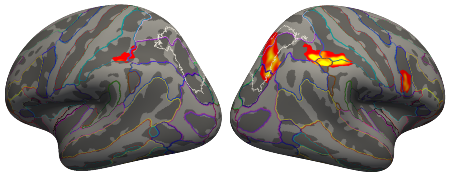


Figure 5.3. Neural activations for ungrouped and grouped stimuli relative to their respective shape control tasks and conjunction analysis. Activation maps obtained from the surface-based group analysis (n = 14) showing the activation elicited by numerosity relative to the shape control task when items in visual arrays were randomly scattered in space (A) or grouped (B). Red and blue clusters respectively indicate greater activity for the numerosity relative to the control task and for the shape control relative to the numerosity task. (C) Brain regions commonly activated by the contrasts shown in (A) and (B). Red clusters represent

the conjunction of the numerosity estimation task performed on both ungrouped and grouped arrays greater than their respective control conditions. Maps are thresholded at $p < 0.001$, uncorrected for multiple comparison and displayed on Freesurfer's fs average surface. Color outlines mark anatomical sulci and gyri according to the Destrieux Atlas (Fischl et al., 2004). White outlines identify the region IPS0-5 based on visual topography (L. Wang et al., 2015). For both contrasts and for the conjunction analysis, the cluster summary tables can be found in A5.

Finally, we looked for regions specifically activated for the number task when the arrays were grouped after subtracting out activity associated with their control tasks, i.e., '(Grouped numerosity > Grouped shape control) > (Ungrouped numerosity > Ungrouped shape control)'. This contrast elicited activations in bilateral angular gyrus and in left frontal regions covering the lateral orbital sulcus, and middle frontal gyrus (red activations in Figure 5.4). On the other hand, no brain regions showed greater activation specifically for the number task when the arrays were ungrouped, i.e., '(ungrouped numerosity > ungrouped shape control) > (Grouped numerosity > Grouped shape control)'.

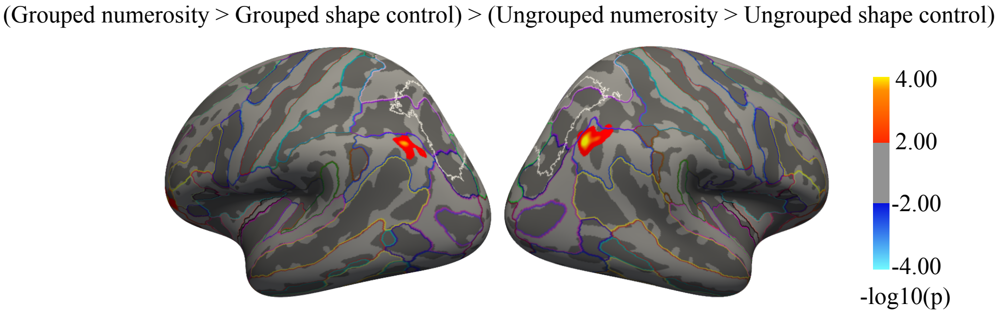


Figure 5.4. Neural activations for grouped compared ungrouped stimuli after subtracting out activity associated with their shape control tasks. Statistical results obtained from the surface-based group analysis showing distinct neural activity for the numerosity task when it was performed on grouped compared to ungrouped arrays. (n = 14). The maps show the regions with greater activations for numerosity estimation of grouped compared to ungrouped arrays, after subtracting out activity associated with their respective shape control tasks. The reverse contrast (grater activations for numerosity estimation of ungrouped compared to grouped arrays, after subtracting out activity associated with their respective shape control tasks) did not yield to significant activation. The cluster summary table can be found in A5.

5.3.4 Multivariate analysis

So far, the results suggest that performing numerosity estimation on ungrouped and grouped arrays elicit partially shared and partially specific cortical areas. However, the fact that similar or overlapping regions are activated by estimation of ungrouped and grouped arrays does not necessarily imply that the same shared neural mechanisms are recruited in both cases, and instead it could reflect functionally different neural patterns within the same brain regions. For this reason, we further evaluated whether the pattern of activity elicited by numerosities with different spatial arrangements showed some similarities or differences across the regions along the dorsal pathway.

We defined five different regions in each participant (Figure 5.5A). Within each region we selected the 1000 most activated voxels to the contrast ('all numerosities > baseline'). These voxels were used to train and test classifiers to discriminate between numerosities within each spatial arrangement (ungrouped and grouped) during the numerosity estimation task. We tested whether the numerosity of a given visual array could be predicted within as well as across each spatial arrangement.

5.3.4.1 Decoding within spatial arrangements

Figure 5.5 B&C showed the performance of the classifier when discriminating between numerosities of ungrouped and grouped arrays respectively. When numerosities were shown with ungrouped arrays the different numerosities could be decoded significantly above chance in all ROIs (V1-V3 $t = 3.65$, $p = 0.001$, $LogBF = 1.5$ V3AB-V7 $t = 3.28$, $p = 0.003$, $LogBF = 1.24$, IPS1-5 $t = 3.92$, $p = 0.0009$, $LogBF = 1.68$, IPS excluding IPS0-5 $t = 3.53$, $p = 0.002$, $LogBF = 1.41$, angular gyrus $t = 3.22$, $p = 0.003$, $LogBF = 1.19$, thus suggesting strong evidence of difference from chance in all ROIs). When numerosities were shown with grouped arrays the decoding accuracy was above chance only in the parietal but not in the early and intermediate ROIs: V1-V3 $t = -0.77$, $p = 0.77$, $LogBF = -0.77$; V3AB-V7 $t = 1.80$, $p = 0.048$, $LogBF = 0.26$; IPS1-5 $t = 3.55$, p = 0.002, LogBF = 1.42; IPS excluding IPS0-5 t = 4.1, p = 0.0006, LogBF = 1.81, angular gyrus t = 2.96, p = 0.006, LogBF = 1.07. Bayes factor (LogBF) provided substantial evidence in favor of no significant difference from chance in the primary ROI, and positive and strong evidence in favor of significant differences in intermediate and parietal ROIs respectively.

Classification accuracies were analyzed with a repeated measures ANOVA with ROIs (5 levels) and spatial arrangements (2 levels) as factors. The results showed a significant interaction between ROIs and spatial arrangements (F(4,52) = 3.702, p = 0.01, p = 0.017 after Greenhouse-Geisser sphericity correction, $\eta^2 = 0.065$). Post-hoc comparisons revealed that the difference in classification accuracy between V1-V3 ungrouped and V1-V3 grouped was at significance (t = 3.53, p = 0.05), and significant for V1-V3 grouped and IPS 1-5 grouped (t = –4.1, p = 0.004) and V1-V3 grouped and IPS excluding IPS 0-5 grouped (t = –4.14, p = 0.003). There was no significant main effect of ROIs and Spatial arrangements (F(4,52) = 2.365, p = 0.065, p = 0.073 after Greenhouse-Geisser sphericity correction, $\eta^2 = 0.062$; F(1,4) = 2.06, p = 0.175, $\eta^2 = 0.042$, respectively for ROIs and Spatial arrangements).

5.3.4.2 Generalization across spatial arrangements

We then tested for generalization of classification performance across spatial arrangements (Figure 5.5D). Significant generalization was observed in the parietal but not in the intermediate ROIs nor in the primary ROIs after correction for multiple comparisons (V1-V3: t = 2.63, p = 0.01, LogBF = 0.79; V3AB-V7: t = 1.51, p = 0.078 LogBF = 0.09; IPS1-5: t = 3.39, p = 0.002, LogBF = 1.31; IPS excluding IPS0-5: t = 4.17, p = 0.0005, LogBF = 1.86; angular gyrus: t = 3.45, p = 0.002, LogBF = 1.35. Bayes factors provided substantial, positive and strong evidence in favor of difference from chance in the primary ROI, intermediate and parietal ROIs respectively).

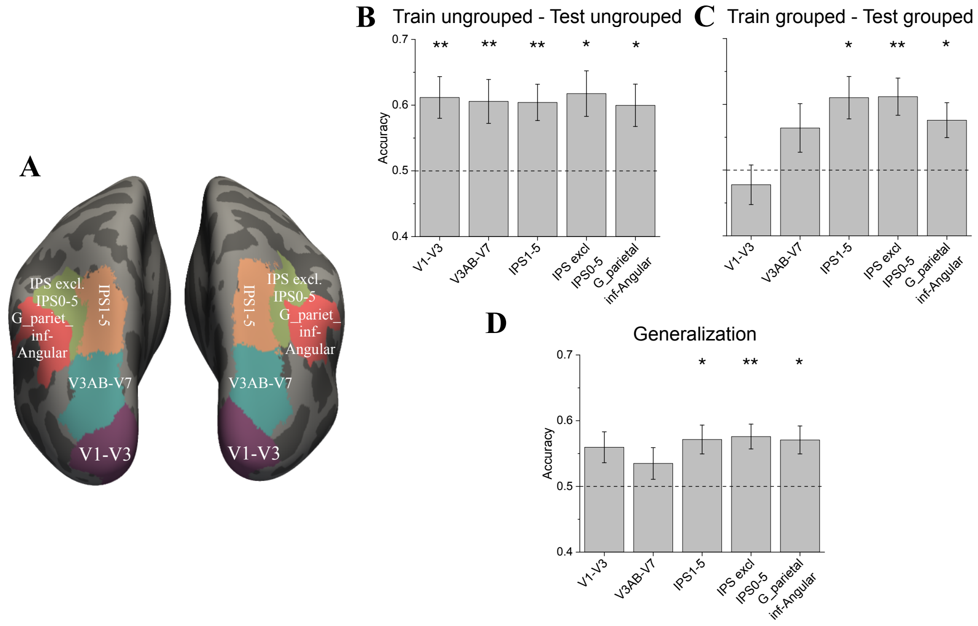


Figure 5.5. ROIs illustration and results of multivariate pattern analysis. ROI localization of the occipital and parietal regions and results of multivariate classification for discrimination between numerosities when participants were performing the numerosity task on items with different spatial arrangement. (A) Color-coded ROIs on the inflated brain template. (B-C) Average decoding accuracy for different numerosities when training and testing the classifier with the pattern of activity elicited by the ungrouped (B) and grouped (C) arrays. (D) Average decoding accuracy for different numerosities when the classifier was trained on the pattern of activity elicited by one given spatial arrangement and tested on the pattern of activity elicited by the other.

Bars show mean classification accuracy across subjects ± standard error of mean (Stars marks significance against chance (0.5), after correction for multiple comparisons: 0.05/5 = 0.01. * p =< 0.01; ** p < 0.001; *** p < 0.0001).

5.3.5 Correlational analysis

We correlated the Weber fraction measured in the behavioral experiment for grouped arrays and the classification accuracy for grouped stimuli based on the neural activity read out from the primary visual areas and the angular gyrus ROIs. As showed in Figure 5.6 classification performance from the angular gyrus correlated significantly with the Weber fraction measured behaviorally ($r = -0.67$, $p = 0.009$, LogBF = 0.86). However, there was no significant correlation with primary visual areas ($r = -0.44$, $p = 0.11$, LogBF = 0.02).

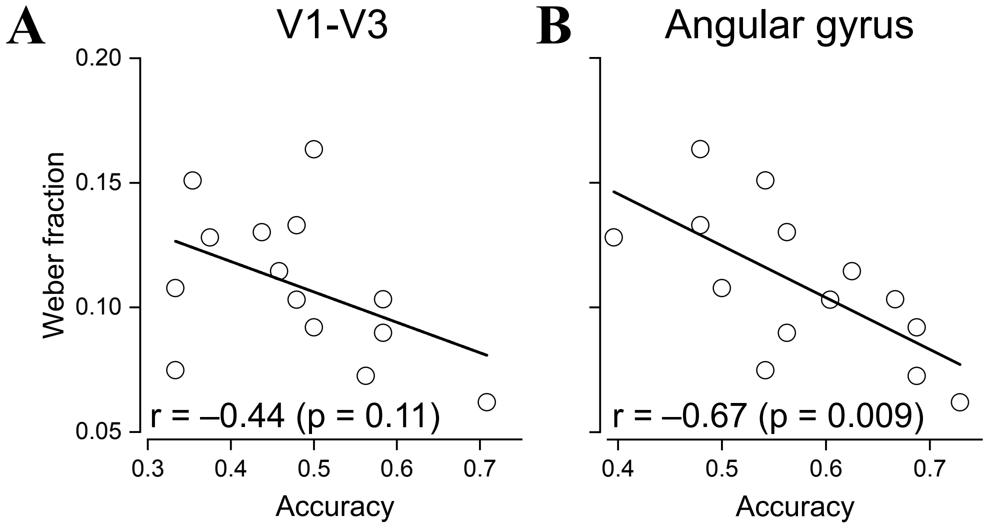


Figure 5.6. Relation between behavioral Wfs and classification accuracy in the numerosity grouped condition. (A) Weber fractions plotted against classification accuracy in the primary visual areas (V1-V3) and (B) in the angular gyrus ROI.

5.4 Discussion

The present study explored the neural resources supporting numerosity estimation when stimuli were grouped (to facilitate "groupitizing") and ungrouped. The results showed that the numerosity estimation of ungrouped and grouped stimuli shared the activation of a similar right lateralized fronto-parietal network. The estimation of grouped stimuli additionally elicited the recruitment of regions in the left hemisphere, specifically the angular gyrus. Multivariate pattern analysis showed that classifiers trained with the pattern of neural activations read out from parietal regions, but not from the primary visual areas, can decode different numerosities both within and across spatial arrangements. Finally, fMRI decoding performance of the angular gyrus but not the primary visual areas correlated with the behavioral Wfs measured in the estimation task.

The results from the univariate analysis of the fMRI data showed that numerosity estimation of ungrouped and grouped stimuli activated both common (mostly right-lateralized) and specific cortical areas within the fronto-parietal network. We identified a right lateralized fronto-parietal circuit that was more activated when participants estimated the numerosity of ungrouped stimuli compared when they estimated the odd-shape in the ungrouped arrays: the circuit including the right superior and transverse occipital sulcus, right IPS, right SPG, right inferior precentral sulcus and superior frontal gyrus and the bilateral postcentral sulcus. We cannot formally rule out the possibility that the activation of the listed areas may in part be driven by levels of different difficulty of the ungrouped numerosity and shape control tasks. However, in the comparison and estimation of non-symbolic tasks the activation of this right lateralized network has been previously reported by neuroimaging studies (Dormal et al., 2010, 2012), even when the difficulty of the numerosity estimation task with respect to its control task was matched (Piazza et al., 2006). Piazza et al. (2006) explored the neural activity associated with estimation of visual and auditory stimuli (squares or tone sequences) and compared it with a control task in which participants were asked to report whether the last stimulus in the sequence was identical to the first. They showed that estimation of these non-symbolic quantities elicited a greater activation of the right IPS, right precentral gyrus and right middle frontal gyrus compared to the control task. Importantly, the main and control tasks in Piazza et al.(2006)'s study were matched for difficulty, therefore this factor could not explain the right-lateralized activity, similar to the one reported in the current study, for numerosity estimation.

In the current experiment, we found that numerosity estimation of grouped stimuli activated similar regions in the right hemisphere but additionally recruited the same areas in the left hemisphere as well. Previous psychophysical evidence suggests that when participants estimate the numerosity of a stimulus divided into subitizable sub-groups they spontaneously employ arithmetical (subitize-and-sum) strategies to solve the task (Maldonado Moscoso et al., 2020; G. S. Starkey & McCandliss, 2014). Evidence in the literature from brain-damaged patients (Cipolotti et al., 1991; Dehaene et al., 1998; Jackson & Warrington, 1986) and fMRI studies (Chochon et al., 1999; Pinel & Dehaene, 2010) have suggested that left hemisphere is involved in mental arithmetic. The major involvement of the left hemisphere may therefore implicate mental arithmetic strategies when items were grouped.

Interestingly, the network recruited for estimating numerosities of grouped arrays is not completely different from that recruited for ungrouped arrays: the conjunction analysis showed that the neural network in the right hemisphere is largely shared between numerosity estimation of grouped and ungrouped arrays. This network includes the right IPS, the right superior and transverse occipital sulcus, the right SPG, the right precentral sulcus and in the right superior frontal gyrus.

Beyond this common shared system, we also observed areas uniquely activated during numerosity estimation of grouped stimuli, specifically the bilateral angular gyrus and in the left orbital sulcus and in the left middle frontal gyrus. Although the angular gyrus has been shown to be activated in various cognitive domains (perceptual and motor reorienting, number processing, attention and spatial cognition, episodic memory retrieval and encoding, language processing, theory of mind; Cabeza et al., 2012), fMRI and neuropsychological studies have shown that the left angular gyrus plays an important role during calculation processing, in particular during multiplication and arithmetical fact retrieval from memory (Chochon et al., 1999; Delazer et al., 2003; Gerstmann, 1940; Grabner et al., 2007, 2013; Grabner, Ansari, et al., 2009; Grabner, Ischebeck, et al., 2009; Ischebeck et al., 2007; Lee, 2000; Stanescu-Cosson et al., 2000; see also: Dehaene et al., 2003, for a review). The activation of the angular gyrus can hardly be explained by task difficulty as on average the accuracy in grouped and ungrouped tasks isn’t statistically different. Rather, it is most likely that this activation reflects calculation procedure as previously reported by other studies (Göbel et al., 2001; Menon et al., 2000; Stanescu-Cosson et al., 2000). Interestingly, the activation of the left inferior and middle frontal regions was found in studies that investigated the neural substrates of symbolic and non-symbolic exact, compared to approximate, calculation. It has been suggested that this left lateralized parieto-frontal network may play an important role for arithmetic fact retrieval (Piazza et al., 2006; Prado et al., 2011, 2014; Stanescu-Cosson et al., 2000). The activity revealed during estimation of grouped stimuli may implicate the automatic use of multiplication strategies or retrieval of arithmetic facts (i.e., basic addition) to solve the task.

Overall, the results from the univariate analysis showed that the numerosity estimation of both ungrouped and grouped stimuli compared to their shape control conditions elicited the activation of similar regions, but with different lateralization, in the fronto-parietal network and with different recruitment of the angular gyrus.

The fact that different tasks elicit the activation of overlapping regions, however, does not necessarily imply that the same neural mechanisms are recruited, but may rather reflect intermingled neural populations that are differentially recruited for the different tasks. We therefore tested whether performing numerosity estimation on arrays with different spatial arrangements elicited similar patterns of neural activation. The results of the multivariate decoding analysis showed that the different numerosities could be read out from brain activity significantly above chance during numerosity estimation of ungrouped arrays all along the visual stream. On the contrary, when training the classifiers with the pattern of activity elicited by estimation of grouped stimuli, numerosity could be decoded above chance only in parietal regions. In this study, the total field area was matched between ungrouped and grouped stimuli, however, within this area, the location of the individual items was more scattered in the ungrouped than in the grouped arrays (by definition). Such difference may have elicited a distinct pattern of activity in the primary visual areas for the ungrouped compared to the grouped stimuli, therefore explaining the higher decoding performance in the former, but not in the latter case.

We also observed that decoding in the parietal, but not the early visual ROIs, successfully generalized across spatial arrangements, suggesting that the numerosity information is similarly encoded in the parietal cortex independently of whether stimuli were ungrouped or grouped.

Finally, we examined the behavioral relevance of the numerosity representations in the early visual areas and angular gyrus by correlating the individual behavioral Wfs against the fMRI decoding accuracies. Our results replicated previous evidence showing that numerical representation did not correlate with behavioral acuity in the primary visual areas (Lasne et al., 2019). However, the correlation between precision of numerosity estimation of grouped stimuli and decoding accuracy in the angular gyrus was significant, with participants with higher precision (lower Wfs) having significantly more accurate decoding of neural activity in this region. The current results may therefore suggest that the angular gyrus plays a crucial role for estimation of grouped stimuli.

In conclusion, we used fMRI to explore neural activation elicited by grouped compared with ungrouped stimuli. Univariate analysis showed that the estimation of both ungrouped and grouped stimuli activate a similar right lateralized fronto-parietal network. The possibility of using grouping strategies may automatically elicit a different strategy for numerosity estimation and results in the recruitment of a network involved in calculation, mostly including regions in the left hemisphere and the angular gyrus. Moreover, that classifiers can generalize across spatial arrangements in the parietal regions suggests that at this level the numerosity information is stored in a more abstract way, which prescinds the spatial arrangement, probably reflecting the participant's response (numerosity estimated). This possibility is supported by the correlation between the participant's behavioral acuity and the decoding performance in the angular gyrus, suggesting that the pattern of activity in this region reflects the perceptual decisions. Overall, and in line with the psychophysical findings, this experiment supports the hypothesis that the estimation of grouped stimuli relies on the system for numerosity estimation, but additionally recruit regions involved in calculation which enable us to more precisely and quickly estimate numerosity of grouped arrays.

Chapter 6

Math anxiety mediates the link between number sense and math achievements in high math anxiety young adults

Chapter 6

6.1 Introduction

Numerical and mathematical competencies are central predictors of an individual's success in life. Developing adequate numerical and mathematical skills is a prerequisite to accomplishing numerous tasks in daily life, such as setting and keeping to a budget (Parsons & Bynner, 2005), as well as pursuing careers in the STEM fields: science, technology, engineering, and mathematics (STEM; Beilock & Maloney, 2015; Ferguson et al., 2015). Impairments in mathematical skills might be triggered by several factors and, amongst these, mathematical anxiety (MA) has been suggested to play a key role. MA has been defined as feelings of apprehension and increased physiological reactivity when individuals have to manipulate numbers, solve mathematical problems, or when they are exposed to an evaluative situation connected to math (Ashcraft, 2002; Hembree, 1990). Similar to other performance-based anxieties, MA involves psychological arousal, negative cognitions, escape and/or avoidance behaviors and, when the individual cannot avoid the situation, performance deficits. MA is also related to reduced cognitive reflection (Morsanyi et al., 2014; Primi et al., 2018), and poorer decision making performance (e.g., Rolison et al., 2016). In other words, MA is described as a multidimensional construct that is related to, but distinct from, other forms of anxiety, such as trait, social, or test anxiety (Ashcraft & Moore, 2009; Vukovic et al., 2013). MA has been shown to hinder math performance. It has been reported that individuals with higher levels of MA obtain lower scores in math achievement tests, take fewer math courses, and tend to avoid career paths involving mathematics (Ashcraft & Krause, 2007; Ashcraft & Moore, 2009; Ma, 1999).

Two theoretical frameworks have traditionally been proposed to account for the link between MA and math achievements (Carey et al., 2016). The deficit theory posits that poor mathematical performance leads to future high levels of MA. In line with that, it has been suggested that MA could result from low numerical (and/or spatial) skills which compromise the development of high proficiency in mathematical problem solving (Maloney, 2016; Maloney et al., 2011). On the other hand, the cognitive interference theory posits that it is MA that affects subsequent mathematical performance. During the phases of information processing and recall, MA would create cognitive interference which affects mathematical performance.

According to this theory, anxiety would generate intrusive thoughts to reduce working memory (WM) capacity, with these thoughts acting as a secondary task draining resources that, otherwise, would have been allocated to solving the mathematical task (Ashcraft & Kirk, 2001). An alternative theory posits that MA and mathematical performance show a bidirectional relationship (Ashcraft & Krause, 2007); past failures and negative experiences in mathematical performance would lead to MA which, subsequently, would lead to poorer mathematical performance and vice versa (Ma & Xu, 2004).

Whatever the nature of the link between MA and low achievement in math learning, several studies have highlighted various factors that might account for the negative relationship between these factors. A possible explanation of the gap in math performance between students with high and low levels of MA derives from behavioral and psychophysiological studies, which provide converging evidence for individual (cognitive, affective/ physiological, motivational) and environmental (social/ contextual) factors (Chang & Beilock, 2016). Recent reports, focused on genetic and neurophysiological factors, suggested that MA arises from a basic level deficiency in symbolic numerical processing. In particular, genetic studies of MA in twins evidenced that genetic factors accounted for about 40% of the variation in MA, and that 12% of the total variance in MA was associated with genetic influences related to math problem-solving (Malanchini et al., 2017; Z. Wang et al., 2014). Finally, children with high mathematical anxiety (HMA), compared with low mathematical anxiety (LMA) peers, show reduced responses in posterior parietal cortex, including the intraparietal sulcus (IPS) and dorsolateral prefrontal cortex regions, known to play a critical role not only in numerical and mathematical cognition, but also in non-symbolic number evaluation (Castaldi et al., 2016; Dehaene et al., 1999; Eger et al., 2003; Piazza et al., 2004; Young et al., 2012).

Whilst symbolic numerical representation and arithmetic are recent cultural inventions specifically adopted by humans, humans share with many non-human animal species an intuitive “approximate number system” (ANS), which is the core ability to automatically and efficiently process numerical magnitude information (Dehaene, 2011). The sensory precision of this system is refined during development and varies considerably between individuals (Halberda et al., 2008, 2012; Odic et al., 2013). It is suggested that numerosity represents a primary visual attribute (Anobile, Cicchini, et al., 2016) and, in line with this idea, recent studies showed that numerosity is spontaneously perceived, even by 5-year old children (Cicchini et al., 2016). Interestingly, several studies reported strong correlations between the precision in numerosity judgments and current, future or past formal mathematical skills in children (Anobile et al., 2013; Anobile, Arrighi, et al., 2018; De Smedt et al., 2009; Feigenson et al., 2013; Halberda et al., 2008; Starr et al., 2013). Complementary studies carried out on subjects with mathematical disabilities (developmental dyscalculia) show that a deficit in mathematical processing generalizes to yield severe difficulties in estimating and comparing numerosity (Anobile, Burr, et al., 2019; Landerl et al., 2004; Mazzocco et al., 2011; Piazza et al., 2010; Pinheiro-Chagas et al., 2014). In light of all these results, some authors suggested that an intact number sense might be a base prerequisite for the later mathematical acquisition or, in other words, that the number sense acted as an early non-symbolic start-up tool for the later development of language-based formal mathematical skills (Butterworth, 1999; Butterworth et al., 2011; Dehaene, 2011; Piazza, 2010).

Given the intimate relationship between MA and mathematical achievements, and the complementary link between these and the ANS, it has also been suggested that there is a possible interplay between ANS and MA. However, evidence collected so far is controversial. In particular, two studies have found that individuals with HMA represent numerical magnitude less precisely than their LMA peers (Maloney et al., 2011; Núñez-Peña & Suaŕez-Pellicioni, 2014). However, as both studies tested with Arabic digits, they only supported a link between MA and symbolic representation of quantity, not numerosity. Recently Braham and Libertus (2018) showed that the association between precision in perceived numerosity (ANS acuity) and subjects' performance in applied problem solving was present only in subjects with HMA levels, suggesting that an efficient ANS system might act as a potential protective factor for highly math anxious students. Another study reported a link between non-symbolic numerical processing and MA (Lindskog et al., 2017); these authors found that people with high levels of math anxiety show poorer precision in a non-symbolic numerical comparisons task, compared to those with low levels of math anxiety. They also showed that the correlation between math skills and numerosity precision was fully mediated by participants' level of MA. However, several studies measuring ANS acuity by means of non-symbolic tasks failed to find a significant correlation between ANS and MA in both adults as well as children, leaving open the question of whether this interplay occurs (Colomé, 2019; Dietrich et al., 2015; Gómez-Velázquez et al., 2015; Hart et al., 2016; Z. Wang et al., 2015).

The current study aims to assess the role of MA in math skills and numerosity perception. We devised two groups with extremely low or high levels of mathematical anxiety (drawn from a large sample of university students) and measured, in both groups, differences in ANS acuity and math abilities as well as correlations between these variables. We first investigated whether the numerosity thresholds were different in subjects with HMA compared to their LMA peers. Then we addressed the question whether any possible numerosity impairments in HMA participants ware selective for numerosity or whether it was related to a more general perceptual weakness in magnitude judgements. This goal was achieved by measuring discrimination thresholds on a non-numerical magnitude task, in which participants were engaged in an object-size discrimination task. The issue of specificity was also tested by measuring a non-magnitude parietal function, as many studies suggested a key role of parietal cortex in both numerosity perception and math processing. To this aim, we decided to administer a Multiple Object Tracking (MOT) task as this task was shown to activate the parietal cortex, which has been found to correlate well with both numerosity and math abilities (Anobile et al., 2013; Ansari et al., 2007; Corbetta & Shulman, 2002; Steele et al., 2012). In order to assess the specific role played by MA in mathematical performance, we measured individuals' anxiety on a more general dimension, such as performance anxiety (Ashcraft & Ridley, 2005; Lindskog et al., 2017). Finally, we tested for the potential mediation role of MA on the link between ANS and math abilities, using a mediation model in which ANS was associated with math achievement through math anxiety. Mediation implies a situation where the effect of the independent variable (*X*) on the dependent variable (*Y*) can be explained using a third mediator variable (*M*) which is caused by the independent variable and is itself a cause for the dependent variable. By modelling an intermediate variable, the overall effect between *X* and *Y* can be decomposed into component parts called the direct effect of X on Y and the indirect effect of X on Y through M (i.e. the mediated effect).

The importance of our study, which took into consideration several possible differences between subjects with high and low math anxiety, relies on the fact that such multidimensional analysis is the most suitable tool to investigate the effect of MA on both low-level quantity processing (ANS) as well as high-level mathematical proficiency. Such an approach is not only likely to allow a full understanding of the interplay between MA, math achievements and ANS, but will also improve understanding of the brain mechanisms underpinning these processes, as well as providing useful information about how to optimize mathematical learning procedures or customized early targeted interventions.

6.2 Methods

Participants were 88 university students attending an introductory statistics course at the School of Psychology of the University of Florence. They were selected from a class of 179 students based on their level of math anxiety. The LMA group comprised 39 participants (69% female; age range 18–22 years, mean = 20.1, SD = 0.7) who scored below the 25th percentile (score range 10-19 , mean = 16.3, SD = 2.6) on the Abbreviated Math Anxiety Scale (AMAS; Hopko et al., 2003). The HMA group comprised 49 participants (82% female; age range 18–37, mean = 20.4, SD = 2.9,) who scored above the 75th percentile on the AMAS (score range 27- 40, mean = 30.1, SD = 3.2). All students participated on a voluntary basis. The whole procedure was performed in accordance with the declaration of Helsinki. Participants were tested individually. Before the testing sessions, students provided informed consent. Math skills (MPP), Math anxiety (AMAS) and Test anxiety (TAI) were all measured before psychophysical experiments. The scales were in a paper-and-pencil format. The psychophysical tasks were then performed in a quiet and dimly illuminated room. Participants sat in front of a BARCO 27” monitor subtending 39° by 29° from the subject’s viewing distance of 57 cm. The monitor resolution was 1024 X 768 and the refresh rate equal to 120 Hz. Stimuli for the psychophysical experiments were all generated and presented with PsychToolbox (Brainard, 1997) routines for MATLAB (ver. 2010a, The Mathworks, Inc.).

6.2.1 Measures

The *Mathematics Prerequisites for Psychometrics* (MPP; Galli et al., 2011) is a test which was developed to measure the mathematical skills of students enrolled in statistics courses. The scale was developed using item response theory (IRT) because it offers a different value of test precision for each specific level of underlying latent variable being measured, and it does not assume that a single estimate of reliability, and corresponding standard error of measurement, is sufficient to describe precision of measurement over all levels of ability (Embretson & Reise, 2000). The scale consists of 30 problems and has a multiple-choice format (one correct response out of four options). For example, "The value 0.05 is" (i) lower than 0; (ii) between −1 and 0; (iii) higher than 0.1; and (iv) between 0 and 1, and "Knowing that xy = 3 which of the following is true?" (i) y=3/x; (ii) y=3x; (iii) c=3x; and (iv) xy/3. The sum of correct responses gave us a single composite score for each participant. In the present sample, Cronbach's α was .73 (IC: .70-.78). We used this measure as an estimate of the students' math knowledge (Primi et al., 2014).

The *Abbreviated Math Anxiety Scale* (AMAS; Hopko et al., 2003) measures MA experienced by students in learning and test situations. Participants were required to respond on the basis of how anxious they would feel during given events (for example, "Listening to another student explain a math formula" or "Starting a new chapter in a math book") by using a 5-point response scale (ranging from strongly agree to strongly disagree). High scores on the scale indicate HMA. A single composite score was obtained, based on participants' ratings of each statement. In the present sample, Cronbach's α was .84 (IC: .80 -.87).

The *Test Anxiety Inventory* (TAI; Spielberger et al., 1978) was developed to measure anxiety associated with task-performing situations in high school and college students. The test consists of 20 items, which investigate a range of anxiety symptoms occurring before, during or after exams. Responses are collected using a 4-point Likert scale ranging from 1 (almost never) to 4 (always). The TAI yields a total score calculated as the sum of all 20 items, with higher scores corresponding to high test anxiety. In the present sample, Cronbach's α was .94 (IC: .93 -.96).

6.2.2 Numerosity discrimination task

Stimuli consisted of two brief (250 ms) patches of dots, presented on either side of a central fixation point (**Figure 6.1A**). Dots were 0.25° in diameter, half white and half black (to balance luminance), presented at 80% contrast on a grey background of 40 cd/m2. They were constrained to fall within a virtual circle of 10° diameter, centered at 10° eccentricity. Standard numerosity (randomly left or right) was fixed at 24 dots while the probe adaptively changed, according to participant responses, with numerosity defined by an adaptive staircase QUEST algorithm (Watson & Pelli, 1983). All participants performed one session of 80 trials. Participants were asked to indicate the side of the screen with more dots. We plotted the proportion of trials where the standard stimulus appeared more numerous than the probe against the probe numerosity (on log axis) and fitted with cumulative Gaussian error functions. We defined the point of subjective equality (PSE) as the physical numerosity of the probe yielding 50% of probe more numerous responses. Then we defined subjects' precision as just notable difference (JND), that is the numerosity offset defining the 50-75% range of probe more numerous. Finally, normalizing PSE by JND we obtained a single index Weber Fraction (WF), a typical dimensionless psychophysical index for discrimination thresholds.

6.2.3 Size discrimination task

Stimuli were gratings sinusoidally modulated in luminance with a spatial frequency of 2 cycles per degree and a Michelson contrast of 90% which were vignetted in an annular contrast window (**Figure 6.1B**). In each trial, two annuli were simultaneously presented for 250 ms on the left and the right side of the central fixation point, at an eccentricity of 10°. Subjects were required to indicate which stimulus appeared to be larger. The diameter of the test stimulus (presented randomly on the left or right) was 5° or 8° (40 trials each, randomized trial-by-trial), while the probe varied in diameter by a percentage drawn randomly from a Gaussian distribution centered at 0 with SD = 20%. To minimize alternative judging strategies (such as estimating border-to-center of the screen distance), we independently jittered the horizontal eccentricity of the test and the probe between 8.5° and 11.5°, and their distance from the horizontal meridian within ± 3°. After the stimuli presentation, a 100 ms full-screen random noise mask was displayed to cancel out possible afterimages. The proportion of "test largest" trials was plotted against the log-ratio of the test to probe and fitted with cumulative Gaussian error functions. Even for the size discrimination task, the dependent variable which we took into account was Weber Fraction (see above), indicating subjects' sensory precision in the size discrimination thresholds.

6.2.4 Visual sustained attention task

Visual sustained attention (**Figure 6.1C**) was measured by a multiple object tracking task (MOT; Pylyshyn & Storm, 1988). At each trial, a total of twelve disks with a diameter of 0.9° moved randomly on the full screen at 7°/s for a period of 2 s. The green targets could be 2, or 3, or 4 (representing the three conditions) and the remaining stimuli (distractors) were red. After the 2 s, the green targets turned red (like the distractors), and continued to move randomly on the full screen for 4 s. The participants were required to continue to track them with their attention. After this period, the disks stop moving, and 4 of them turned orange. Participants had to identify (using the mouse cursor) which one of the four orange items was a green target at the beginning of the trial (4AFC). Each experimental session had 10 trials and participants performed 2 sessions, for a total of 20 trials. No feedback was provided. We measured the performance of the participants as the proportion of correct responses for each condition (Anobile et al., 2013).

6.2.5 Data analysis

Preliminarily, we tested differences within the group (LMA and HMA) on numerosity and size discrimination tasks as well as sustained attention with a mixed 3 (within factor: tasks) X 2 (between factor: groups) ANOVA. Correlations between variables were tested by Pearson's r. To further enhance the understanding of the mechanisms underlying the relationships among these variables, a mediation model was tested. Specifically, MA was modelled as the intermediate variable (M) between ANS and math proficiency. This procedure allowed us to conclude whether the independent variable influences the dependent variable directly (path c' in **Figure 6.5**) and/or indirectly (path a or b in **Figure 6.5**) through the mediator. Obviously, the direct and indirect effects added to the yield of the total effect (path c in Figure 5) of the independent variable on the dependent variable. The mediation model was estimated to derive from the total, direct, and indirect effects of ANS on math achievement through MA. The indirect effect of ANS on math achievement was quantified as the product of the ordinary least squares (OLS) regression coefficient estimating MA from ANS (i.e., path a in **Figure 6.5**) and the OLS regression coefficient estimating math achievement from MA when controlling for ANS (i.e., path b in **Figure 6.5**). To test the mediation model, we used the INDIRECT macro for SPSS (Hayes, 2013). The INDIRECT macro tested the hypothesized model using a bootstrapping procedure (with 5000 bootstrap samples) to estimate the 95% confidence interval for the indirect (mediated) effect (for more details, see Preacher & Hayes, 2008). Bootstrapping is a resampling strategy for estimation and hypothesis testing. With the bootstrapping method, the sample is conceptualized as a pseudo-population that represents the broader population from which the sample was derived, and the sampling distribution of any statistic can be generated by calculating the statistic of interest in multiple resamples from the dataset. The bootstrapping procedure has been suggested as representing the most trustworthy test for assessing the effects of mediation models, overcoming issues associated with inaccurate p-values which result from violations of parametric assumptions (Hayes & Scharkow, 2013). Indeed, the bootstrapping procedure is advantageous because it does not impose the assumption of normality on the sampling distribution of indirect effects, and it retains high power while maintaining adequate control over Type I error rate (Hayes, 2009; MacKinnon et al., 2002, 2004; Preacher & Hayes, 2008). The bootstrap test is statistically significant (at .05) if both confident limits have the same sign (e.g., both positive and both negative). This indicates that zero is not a likely value, and therefore, that the null hypothesis of a null indirect effect has to be rejected.

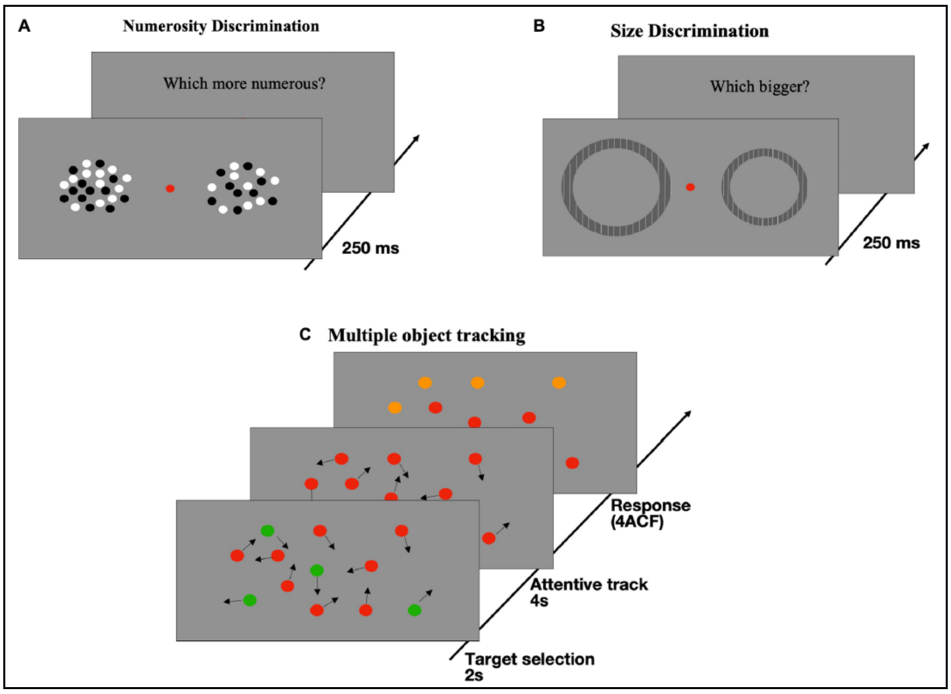


Figure 6.1. Stimuli and procedure. (A) Numerosity Discrimination: two patches of dots were briefly (250 ms) presented to both side of a central fixation point. Subjects were required to select which dots ensemble was more numerous. (B) Size Discrimination: Participants were asked to indicate which of two briefly (250 ms) presented annuli was perceived as being larger (method adapted from Pooresmaeili et al., 2013). (C) Multiple Object Tracking (MOT): At the beginning of the session, some disks (2, 3, or 4) out of twelve were colored in green with the remaining being red. All dots moved randomly on the screen (7°/s) for a period of 2 s then the green disks turned red (like the distracters) and subjects had to track them for 4 s. At the end of the tracking period, all dots stopped and 4 of them turned orange with one

of the orange dots being green at the beginning. This dot was the target subjects had to indicate in a 4-alternative forced paradigm (4ACF).

	LMA			HMA		
	M	SD	N	M	SD	N
ANS Wf (%)	23.57	8.69	39	24.41	9.02	49
Size Wf (%)	12.15	8.26	39	9.96	4.27	49
Attentional Index	0.69	0.11	39	0.71	0.1	49
Math performance	23.63	3.51	38	21.33	3.75	49
Math anxiety	16.36	2.57	39	30.08	3.18	49
Test anxiety	34.34	9.42	38	55.20	12.12	49

Table 6.1. Descriptive statistics for LMA and HMA groups

6.3 Results

6.3.1 Differences between groups

At first, we measured the difference in math anxiety between the students in the HMA and LMA group that turned out in being highly statistically significant (t(86) = – 21.85, p < 0.001). We then measured performance difference between HMA and LMA groups in the psychophysical tasks (see **Table 6.1** for descriptive statistics). Numerosity and size discrimination thresholds (WF) were measured separately for each participant. Attentional performance in the MOT task was computed as a percentage of correct responses separately for the three experimental conditions (tracking of 2, 3 or 4 dots) however, given all these conditions turned out to be highly correlated to each other (Mot 2 and Mot 3 r = 0.351, p < .001; Mot 2 and Mot 4 r = 0.305, p = .004; Mot 3 and Mot 4 r = 0.61, p < .0001), we computed a single index to estimate the performance in the attentional task by averaging the scores across conditions. Individuals in the low and high math-anxiety groups, showed similar performance across all tasks (F(1,86) = 0.036, p = 0.85); the interaction was also not significant (F(2,172) = 1.539, p = 0.218). Post-hoc t-test confirmed the differences between groups were not significant in both, numerosity and size discrimination tasks (Numerosity Wf: t(86) = -0.444, p = 0.658; Size Wf: t(86) = 1.607 p = 0.112, **Figure 6.2A and 6.2B**). Similarly, performance in the attentional task did not turn out to be statistically significant between the two groups considering neither the aggregate index, nor each experimental condition (defined by the number of objects to track) independently (Mot 2: t(86) = -0.24 p = .8; Mot 3: t(86) = -1.95 p = .05; Mot 4: t(86) = 0.28 p = .78). Finally, not only the LMA group had statistically higher math proficiency but also lower test anxiety scores compared to the HMA group (t(85) = 2.923, p = 0.004; t(85) = -8.75, p < 0.001 for math performance and test anxiety score respectively).

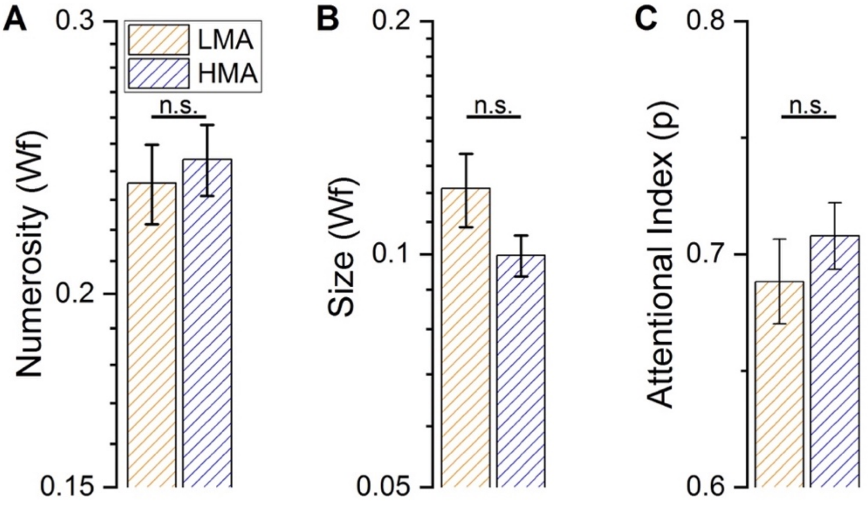


Figure 6.2. Performance in the three different psychophysical tasks. (A) Average numerosity discrimination thresholds (Weber fraction) for subjects with high (HMA) and low (LMA) levels of math anxiety. (B) Average object-size discrimination thresholds (Weber fraction) for subjects with high (HMA) and low (LMA) levels of math anxiety. (C) Average proportion of correct response in the Multiple Object Tracking task, for subjects with high (HMA) and low (LMA) levels of math anxiety.

6.3.2 Correlations between variables

After showing that the two math-anxiety groups did not differ in their precision to discriminate stimuli numerosity or size and were also comparable in terms of attentional performance, we investigated the relationships between perceptual and non-perceptual measures within the two groups (see **Table 6.2** for full correlation values).

Measure	1	2	3	4	5	6
1. Math performance	-	-.290*	-.186	-.014	-.479***	-.009
2. ANS acuity	-.205	-	-.062	-.082	.481***	.073
3. Size acuity	-.139	-.023	-	.128	-.065	-.156
4. Attentional index	.242	-.330*	-.297*	-	-.255*	-.212
5. Math anxiety	-.261	.073	.140	-.256	-	.104
6. Test anxiety	.087	-.008	-.108	.047	.072	-

Table 6.2. Pearson correlation. Pearson correlations between all measured variables in the HMA sub-group (above diagonal) and LMA sub-group (below diagonal). * $p < 0.05$, *** $p < 0.0001$

For clarity, we will describe the data separately for the two math-anxiety groups. Within the HMA group, results demonstrated a significant correlation between MA level and math abilities, with individuals with higher levels of MA having lower math scores ($r = -0.479$, $p < 0.001$; **Figure 6.3**).

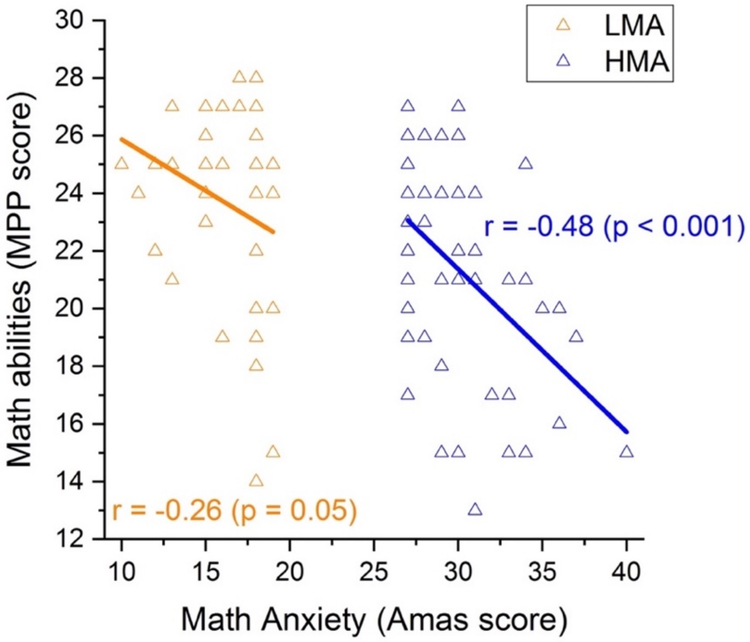


Figure 6.3. Correlation between math anxiety and math abilities. Correlations between math anxiety and math in participants with LMA (orange) and those with HMA (blue).

Moreover, participants with worse numerosity thresholds (higher Wf) also showed higher levels of MA ($r = 0.48$, $p < 0.001$; **Figure 6.4A**) and lower math scores ($r = –0.29$, $p < 0.02$; **Figure 6.4B**). Interestingly, object size discrimination thresholds were not related to math anxiety level ($r = –0.065$, $p = 0.33$, see **Table 6.1**) nor to math scores ($r = –0.19$, $p = 0.1$, see **Table 6.2**). Within the HMA group, participants with better performance in the Multiple Object Tracking task (MOT) also had lower math anxiety levels ($r = –0.255$, $p = 0.04$, see **Table 6.2**). All the remaining correlations with the MOT task were not statistically significant ($p > 0.05$). Finally, test anxiety did not significantly correlate with any of the aforesaid variables ($p > 0.05$, see **Table 6.2**). To further assess the specificity of the link between ANS, MA and math scores, we ran a series of partial correlations taking into account, as covariates, size acuity (WF) and attentional performance (attentional index). These analyses were only run within the HMA group, where bivariate correlations turned out to be statistically significant coefficients. Results of partial correlations revealed that the link between ANS acuity and math anxiety, as well as with math performance, remained statistically significant even when simultaneously controlling for the effects of size acuity, attentional performance and test anxiety ($r_{(partial)}=0.478$, $p<0.001$, $r_{(partial)}= -0.3$, $p=0.019$ for math anxiety and math performance respectively).

Within the LMA group, the pattern of correlations changed significantly. Despite math anxiety and math abilities being (marginally) negatively correlated ($r = -0.26$, $p = 0.05$; **Figure 6.3**) within this group, numerosity discrimination thresholds were not related to math-anxiety levels ($r = 0.07$, $p = 0.33$; **Figure 6.4A**) nor to math scores ($r = -0.20$, $p = 0.1$; **Figure 6.4B**).

In order to check whether the lack of correlations between numerosity thresholds and MA, and math scores in the group with LMA was due to a difference between subject variance for WF between High and Low anxious individuals, we analyzed and compared variance of numerosity thresholds in the LMA and HMA groups by means of a bootstrap technique (Anobile, Arrighi, et al., 2019). On each of 10,000 iterations (sample-with-replacement), we computed Wf average standard deviation in the LMA and HMA groups separately. We then statistically computed the difference between HMA and LMA by counting the number of times that, in each of the 10,000 iterations, the difference between the average in the HMA sample was higher than the average in the LMA sample (one-tailed p value). The p-value was 0.56, suggesting that the lack of correlations described above did not depend on a different level of variance in the data of the two (LMA and HMA) groups. With the same procedure we also excluded a difference in the degree of variability in the MA scores between the two groups ($p = 0.1$)

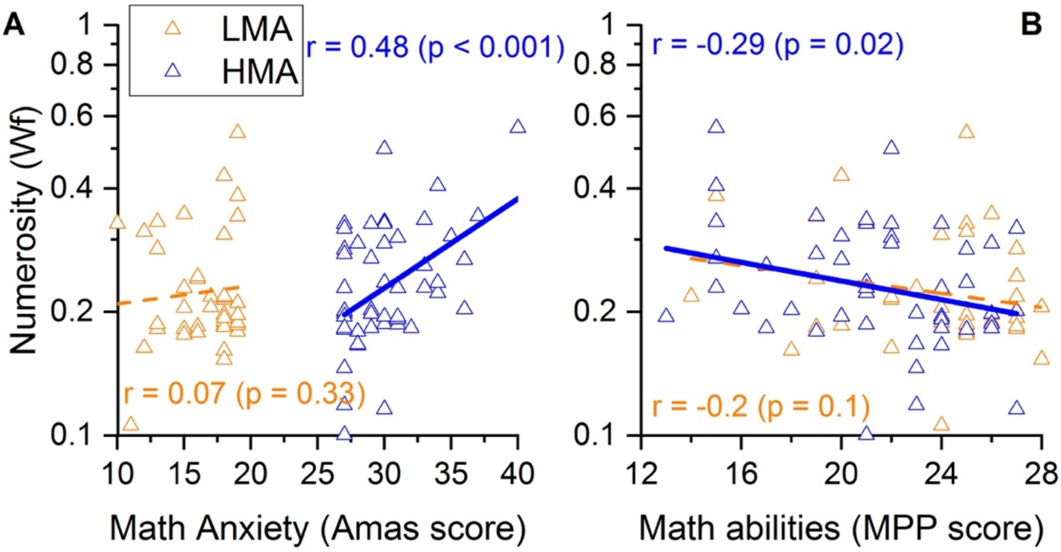


Figure 6.4. Relationship between ANS, MA and math abilities. (A) Correlations between Numerosity discrimination thresholds and math anxiety or (B) math scores for the low math anxiety participants (in orange) and high math anxiety participants (in blue).

6.3.3 Mediation analysis

Given the robust link between numerosity perception (ANS) and math abilities in the group with HMA (see right panel in **Figure 6.4**), we explored the nature of this link by measuring the mediating role of MA. For this purpose, we ran a mediation model to derive the total, direct, and indirect effects of ANS on math achievement through MA. As shown in **Figure 6.5**, results indicate a significant total effect of ANS on math achievement while the direct effect, their relationship not mediated by MA, was found to be not significant. In contrast, a significant negative indirect effect of ANS on math achievement was found when MA was considered as a mediator. Indeed, the bias-corrected bootstrap 95% CI for the product of these paths (a-b) did not include zero (point estimate = -0.08, 95% CI = [-0.1459, -0.0109]), indicating an indirect effect (Preacher & Kelley, 2011).

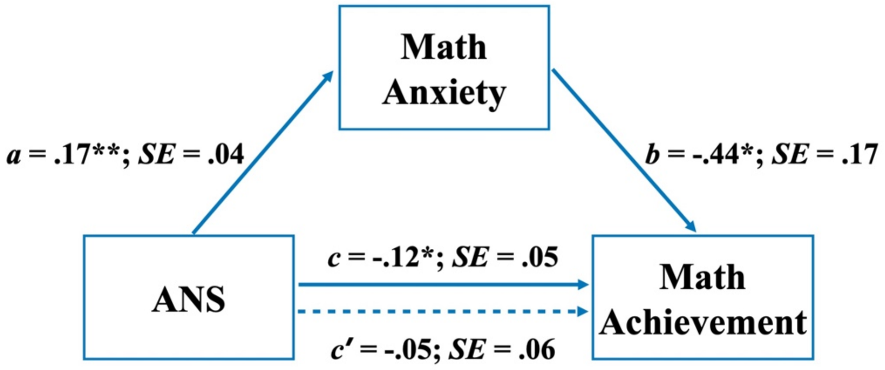


Figure 6.5. Mediation analysis. Path coefficients for mediation analysis on achievement; a, b, c, and c' are unstandardized ordinary least squares (OLS) regression coefficients. *p < 0.05; **p < 0.01.

6.4 Discussion

In the current study, we found that numerosity and object size discrimination thresholds, as well as the ability to attentively track objects in space (MOT), did not differ, on average, between university students with high and low levels of math anxiety. Interestingly, within the high math-anxiety group, numerosity (but not object size) thresholds correlated with both math abilities scores and math-anxiety levels. Crucially, the link between numerosity and math was fully mediated by math-anxiety levels. Overall, our data replicates previous studies on the link between math abilities and numerosity perception but also provided innovative information on the key role that math anxiety plays in such a relationship. Moreover, the fact that math anxiety was found not to be related to size discrimination thresholds, nor to the ability to attentively track objects in space (MOT), strongly suggests that the link between numerosity perception and math-anxiety is not generic but reflects a specific relationship within the numerosity-domain.

Several previous studies have shown that individuals with HMA performed worse on several numerical and mathematical tasks, compared with their low math anxious peers (Ashcraft & Faust, 1994; Maloney et al., 2011). Individuals with lower levels of mathematical skills and high levels of math anxiety show the tendency to avoid situations and careers that require mathematical abilities (Ashcraft, 2002; Hembree, 1990). Given the significant impact of MA on an individual's quality of life, it is important to better understand its nature. Moreover, to devise successful supporting strategies to reduce the level of anxiety related to math procedures, it might be important to find a predictor or a correlated dimension to MA which could be assessed even before the beginning of school. Some studies suggest that such a dimension might be ANS acuity.

In the current study, we tackled this issue by investigating whether the performance in several perceptual tasks concerning parietal driven magnitude processing (discrimination of stimuli numerosity or size) were related to MA as well as math proficiency. We found that MA is an intermediary factor in the link between math abilities and numerosity perception (ANS acuity) in individuals with HMA. The ANS is considered to have evolutionary roots and it appears very early during development (Dehaene et al., 1998; P. Starkey et al., 1990). Maloney et al. (2010) suggested that a deficit of basic and core numerical knowledge, such as numerical information, could produce MA. By taking into account individuals located in the tails of the MA distribution, a procedure exploited by several previous studies (Colomé, 2019; Maloney et al., 2010, 2011; Núñez-Peña & Suaŕez-Pellicioni, 2015; Suaŕez-Pellicioni et al., 2013), and by considering as a measure of ANS acuity the Weber Fractions (Wf; Halberda et al., 2008; Mazzocco et al., 2011; Piazza et al., 2004, 2010), we found that a significant correlation between ANS precision and MA only exists in HMA groups. Our data shows that individuals with very high levels of MA also have a noisy approximate number sense. Notably, the lack of correlation in the LMA group between these two variables was not due to a difference in variability between the two samples. These results are not just important per sè, but also because they are likely to resolve the controversy in the literature about a possible link between MA and ANS precision. For example, Lindskog and coll. (2017) reported that individuals with high levels of MA also show lower ANS precision compared to low mathematics-anxious individuals. However, other studies reported that MA and ANS acuity did not significantly covary in adults (Braham & Libertus, 2018; Dietrich et al., 2015) or in children

(Hart et al., 2016; Z. Wang et al., 2015). One possibility is that MA and ANS acuity covaried differently according to the MA level. For example, in the present study a significant correlation between these two dimensions was found just within the group of participants with HMA. On the contrary, by considering all participants as a whole, MA and ANS acuity shows a weaker correlation that turned out to be marginally significant. In other words, ANS precision and MA strongly correlated in the group of HMA individuals but much less in the group of LMA. If so, the statistical significance of the correlation amongst these dimensions, when the two groups are not independently taken into account, depends on the amount of HMA participants and the severity of their anxious levels, variables which robustly differed in the studies reporting conflicting results in the literature.

Our data highlighted another important point: individuals situated at the lower tail of the HMA group performed better in the numerosity task than the individuals situated in the upper tail of the LMA group. This result supports the idea that an “optimum” level of MA might exist which, if exceeded, becomes deleterious not only for math performance (Evans, 2002), but also for discrimination of abstract numerosity. Furthermore, our findings provide supporting evidence for the theory that individuals with a noisy approximate number system may be more likely to have significant levels of MA. Poor ANS could increase the probability of going through an initial failure and negative learning experience during math education in childhood (Lindskog et al., 2017). One possible explanation of our data is that math abilities and ANS (Weber fraction) are separate (partially independent) predictors of MA, suggesting a bidirectional relationship between MA and math performance, in which a poor ANS induces a low performance in math related tasks and this, in turn, induces MA. This increase in MA might, subsequently, negatively impact math performance, establishing a vicious cycle that dramatically affect an individual’s performance and quality of life.

Math anxiety is strongly correlated with math abilities in individuals with HMA. In line with previous studies, we found that higher levels of MA are linked to lower performance in school or college tests (Hembree, 1990; Ma & Kishor, 1997). MA is at least partly related to fear of failure, so that repeated experiences of failure in mathematics, involving low scores in formal assessments or personal experience of confusion and bewilderment in mathematical activities, may lead to anxiety. Our results are also in line with other studies showing that adults with higher precision in discriminating non-symbolic quantities show higher abilities in math performance (Braham & Libertus, 2018; Fazio et al., 2014; Libertus et al., 2012; Lindskog et al., 2017; Schneider et al., 2017). However, it should be mentioned that, despite many studies which found statistically significant correlations between math abilities and numerosity perception, the literature on this topic is still controversial as other studies report insignificant correlations (Inglis et al., 2011; Krueger, 1984) and the direction of the causal link between ANS and mathematical skills remains highly unclear. While some research suggests that the ANS is a precursor of later mathematical abilities (Anobile et al., 2013; Gilmore et al., 2010; Park & Brannon, 2013; Piazza, 2010) other research failed to find a correlation between ANS precision and mathematical achievements (Anobile, Arrighi, et al., 2018; Feigenson et al., 2013; Inglis et al., 2011; Krueger, 1984). Even if the reasons subtending these discrepancies are still unclear, recent works suggested the important role of the different tests used to assess formal math abilities (Anobile et al., 2013; Braham & Libertus, 2018; Lourenco et al., 2012; Piazza et al., 2010) the numerical ranges used to assess numerosity perception (Anobile, Arrighi, et al., 2019; Anobile, Castaldi, et al., 2016) as well as the age of the participants (Anobile, Arrighi, et al., 2018; Inglis et al., 2011). For example, Braham and Libertus (2018) recently found that students' ANS acuity did not correlate with their ability to perform mathematical computations in written format, but the correlation occurred with their ability to perform speeded mental arithmetic and quantitative reasoning problems. Similarly, Anobile et al. (2013) found that numerosity thresholds in neurotypical primary school children were related to math tasks requiring the encoding of digit magnitude (e.g., choose the largest among others) but not with those more related to memory (e.g., tables) or transcoding (e.g., number writing or repetition), replicating evidence on dyscalculic children (Piazza et al., 2010). Other recent works suggested that the link between numerosity perception and math is present only for the perception of intermediate numerosity levels and not for very low (Anobile, Arrighi, et al., 2019) or very high (Anobile, Castaldi, et al., 2016) numerous ensembles. The current study makes the general picture even more complicated as we found a significant correlation between math and ANS only among adults with relatively high level of math anxiety.

The mathematical test used in the current study, which was developed by Galli and coll. (2011), includes 30 multiple-choice questions covering many aspects of arithmetic knowledge, such as probabilistic reasoning, use of fractions, percentages, ratios, calculation, sorting and others. The test, as a whole, is capable of differentiating subjects with low and high MA and also correlates with numerosity thresholds, at least in the high anxiety group. Future studies on larger and more heterogeneous populations than that involved here, could analyse if and which of these 30 items are more specifically related to both anxiety and numerosity perception.

In addition to the controversial literature on the link between numerosity perception and math abilities, an influential recent theory challenged the idea that numerosity can be encoded by a specialized numerical system. This theory suggests that numerosity and other continuous quantities, such as objects sizes, are perceived by a generalized magnitude system (Henik et al., 2017; Leibovich et al., 2017). In the present study we didn’t find a significant correlation between size and numerosity threshold (Weber fractions). Moreover, whilst numerosity WFs were found to be significantly correlated with math scores, the correlation between math performance and size discrimination thresholds turned out in being not significant. These results clearly contradict the generalized magnitude theory and agree with studies suggesting separate mechanisms for the perception of objects’ numerosity and size. Among these, a recent study found similar results, with no correlations between numerosity and size thresholds as well as between numerosity and size sensory adaptation magnitudes, in both children and adults (Anobile, Burr, et al., 2018). Regarding the selective link between numerosity and math abilities, Piazza, et al. (2013) showed that the exposure of non-schooled indigenous peoples to mathematical knowledge improves the sensitivity to numerosity but not to the size of objects. Similarly, Anobile, Burr, et al. (2018) found that discrimination thresholds for numerosity, but not for objects size, is compromised in dyscalculia.

Overall, despite being still under debate, our results favour the idea of a specialized numerosity system, specifically linked to math abilities and math anxiety.

We didn't observe an impairment in the performance of the visual sustained attention task in subjects with HMA, suggesting that they don't suffer from a general attentional problem despite previous studies in the literature reporting that sustained attention correlates with non-symbolic numerical perception and mathematical skills (Anobile et al., 2013; Steele et al., 2012). Taken together, these results suggest that the link between non-symbolic numerical processing and MA is genuine and does not arise from a generic deficit in the processing of magnitude information or a generic attentional deficit. Even though our approach did not allow us to infer causal connections between the variables we investigated, and the present results cannot be generalized due to the specific sample we chose (students from the Psychology school with un unbalance sampling between male (34%) and female (76%) students), our findings might have important implications in the study of the relationship between ANS and mathematical skills in children with and without mathematical difficulties (e.g. dyscalculia), where MA is meant to play a key role. Indeed, the present results make clear that, in addressing deficits in mathematical performance, low-level aspects such as the ANS acuity as well as high-level aspects as MA have both to be considered. Future research may test the role of MA in the relationship between ANS and mathematical skills in a population of school-age children with a typical development as well as in age-matched subjects affected by dyscalculia, information which would provide a more detailed description of the interplay between MA, ANS and math proficiency.

Chapter 7

General discussion

Chapter 7

7.1 Overview of the findings

Several issues were tackled in this thesis such as the adaptability of non-symbolic processing using novel techniques; the role of stimuli configuration in defining subjects' estimation ability, the nature of the relationship between grouped/ungrouped stimuli and calculation abilities, the different neural substrate of the estimation of grouped and ungrouped stimuli and, eventually, the mutual relationship between the ANS, math abilities and math anxiety. I first reported an experiment aimed to investigate the differences in terms of the shift of minimal confidence and maximal response time relative to the shift of point of subjective equality as a consequence of either visual or motor numerosity adaptation. The results showed that both kinds of adaptation changed the stimulus intensity value providing the maximum uncertainty. Similarly, also the peak for the slowest response times occurred at the adapted point of subjective (rather than physical) equality of the matching task, suggesting that adaptation (both perceptual and motor) acts directly on the sensory representation of numerosity, before any decisional process.

I then focused on extending recent results about the groupitizing phenomenon. First, we demonstrated that groupitizing is able to shoot down the sensory noise of numerosity estimation, probably by triggering implicit and adaptive mathematical strategies. Then, we devised an experiment to test several key assumptions of groupitizing by leveraging on an attentional manipulation (known to strongly affect subitizing) and correlating the groupitizing advantage, as well the groupitizing thresholds with calculation skills. We found that, when subjects got engaged in a concurrent task while performing numerosity estimation (dual task paradigm), showed a strong impairment in the estimation precision of grouped, but not of ungrouped arrays. In other words, the results suggest that groupitizing (like subitizing) is an attention-based process that leverages on subitizing system and calculation abilities. What is more, measuring numerical estimation thresholds with grouped rather than ungrouped items may provide a more sensitive and robust index of math abilities in adults. Strong evidence in support of the mechanism underlying the groupitizing phenomenon comes from the fMRI experiment. Performing the estimation of grouped compared to ungrouped stimuli elicited the activation of overlapping regions in the right fronto-parietal network, and the additional activation of the left fronto-parietal network typically involved during calculation tasks. Specifically, the involvement of the angular gyrus suggested the retrieve of arithmetic facts (i.e., basic additions or multiplications) in groupitizing strategies, supporting behavioral results.

Second, the measurement of ANS precision of ungrouped stimuli in a cohort of university students with either high or low math anxiety while also assessing a) math proficiency using a standardized test; b) visuo-spatial attention capacity by means of a Multiple Objects Tracking task, and c) the sensory precision for non-numerical quantities. The results show a significant correlation between math abilities and ANS precision in participants with high math anxiety. The data also revealed a relationship between ANS precision and math anxiety, and a mediation analysis revealed the mediator role of math anxiety in the relationship between ANS and math abilities showing a genuine interplay between extreme levels of math anxiety and the sensory precision in the processing of the non-symbolic numerosity. These results open up to the possibility of early interventions for subjects showing low ANS acuity that might prevent the development of math anxiety and promote successful math learning.

7.2 Conclusion

Overall, the studies presented in the present work explored different aspects of the complex and multifaceted field of numerosity perception.

The adaptation technique has been used in an extensive portion of literature as a tool to prove the existence of a dedicated numerosity perceptual mechanism. Interestingly, the results from the first study presented in this thesis provides pivotal evidence to this field. Indeed, in the study on visual and motor adaptation, we validated Gallagher et al.'s (2019) pioneering technique by measuring changes in perceptual numerosity, confidence level and response time. The first result of the study is that visual numerosity adaptation effects not only numerosity judgments but also subject's confidence and reaction times. More interestingly, the study demonstrated that motor adaptation changes the numerosity experience at a perceptual rather than at a cognitive/decisional stage. This study linked with previous evidences has brought to hypothesize the existence of a sensorimotor numerosity system, which encodes both external stimuli and internally generated actions. This sensory-motor system seems to be part of a wider generalized system interfacing action with the processing of space, time and number magnitudes (see: Anobile, Arrighi, et al., 2020). The link between action and numerosity opens up new questions: for instance, about their reciprocal interaction in the typical and atypical development and the role of motor skills and the ANS in the acquisition of the symbolic mathematical knowledge.

The relationship between numerosity perception and math skills is important for its consequences in a wide number of fields, such as diagnostic and rehabilitation of mathematical skills impairments. Many studies have suggested that some aspects of estimation of numerosity is related to math. However, the evidence is non-conclusive, as mentioned above in this thesis, with some studies reporting a relationship between numerosity perception precision and formal arithmetical learning during the development while in educated adults this association appears to be less evident. The studies described in this thesis suggest that a promising candidate for the link between ANS and math abilities could be "groupitizing: the strategy of grouping parse arrays into subitizable chunks. This strategy requires some basic arithmetical skills: simple but rapid addition or multiplication. In support to this hypothesis I reported three experimental studies investigating the underling mechanism of groupitizing. Even though in Chapter 3 we didn't measure directly the link between estimation thresholds and math abilities we still found a relationship between subjective precision in the estimation of ungrouped and grouped configurations, suggesting that grouping strategies were used also in the ungrouped patterns by participants with a better ANS, that allowed them to take advantage of intrinsic grouping that could be found even in ungrouped configurations. Adults' ability to take advantage of grouping strategy may be related to their mathematical abilities: for this reason, in Chapter 4 we step forward into the analyses of groupitizing strategies and we found that those participants who took more advantage from grouped configuration also had better arithmetical abilities. The fMRI study indeed confirmed a neural basis for the use of calculation strategies (such as multiplication or addition) for grouped stimuli.

Interestingly, in all of the studies we found that some subjects spontaneously (and also adaptively) use the spatial configuration of stimuli to boost their performance in an estimation task, this might imply that these strategies are also implemented during the estimation of ungrouped stimuli as well. It is likely that the controversial results in the relationship between ANS and math abilities could be related to the variability in strategies used by participants during estimation tasks. The reasons why some healthy adults are not able to take advantage of the explicit grouping isn't explored in the present thesis, however it is a crucial aspect to investigate in future studies.

To conclude, in Chapter 6 I presented a study suggesting fascinating implications to the field of numerosity perception. In fact, we found a link between math and ANS only among adults with a relatively high level of math anxiety. A likely explanation of this result is that individuals with a poor ANS performance show worse results in mathematical tasks compared to individuals with a normal or high performing ANS, and this increases their susceptibility to having math anxiety.

In conclusion, in light of the presented results, ANS precision could lose part of its predictive power of the mathematical skills of adults as it has been refined and mapped onto symbolic numbers. In correlational studies on adults, the lack of a relationship between ANS and math abilities should be taken with caution since, as proved here, behind this relationship there are a lot of perceptual, cognitive, and affective variables that could influence this relationship.

References

Aagten-Murphy, D., & Burr, D. (2016). Adaptation to numerosity requires only brief exposures, and is determined by number of events, not exposure duration. *Journal of Vision*, *16*(10). https://doi.org/10.1167/16.10.22

Agrillo, C., Dadda, M., Serena, G., & Bisazza, A. (2008). Do fish count? Spontaneous discrimination of quantity in female mosquitofish. *Animal Cognition*, *11*(3). https://doi.org/10.1007/s10071-008-0140-9

Agrillo, C., Dadda, M., Serena, G., & Bisazza, A. (2009). Use of number by fish. *PLoS ONE*, *4*(3). https://doi.org/10.1371/journal.pone.0004786

Alais, D., & Burr, D. (2004). Ventriloquist Effect Results from Near-Optimal Bimodal Integration. *Current Biology*. https://doi.org/10.1016/S0960-9822(04)00043-0

Alexi, J., Cleary, D., Dommisse, K., Palermo, R., Kloth, N., Burr, D., & Bell, J. (2018). Past visual experiences weigh in on body size estimation. *Scientific Reports*. https://doi.org/10.1038/s41598-017-18418-3

Allïk, J., & Tuulmets, T. (1991). Occupancy model of perceived numerosity. *Perception & Psychophysics*, *49*(4). https://doi.org/10.3758/BF03205986

Anobile, G., Arrighi, R., & Burr, D. C. (2019). Simultaneous and sequential subitizing are separate systems, and neither predicts math abilities. *Journal of Experimental Child Psychology*. https://doi.org/10.1016/j.jecp.2018.09.017

Anobile, G., Arrighi, R., Castaldi, E., & Burr, D. C. (2020). A Sensorimotor Numerosity System. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2020.10.009

Anobile, G., Arrighi, R., Castaldi, E., Grassi, E., Pedonese, L., Moscoso, P. A. M., & Burr, D. C. (2018). Spatial but not temporal numerosity thresholds correlate with formal math skills in children. *Developmental Psychology*. https://doi.org/10.1037/dev0000448

Anobile, G., Arrighi, R., Togoli, I., & Burr, D. C. (2016). A shared numerical representation for action and perception. *ELife*, *5*. https://doi.org/10.7554/eLife.16161

Anobile, G., Burr, D. C., Gasperini, F., & Cicchini, G. M. (2019). Near optimal encoding of numerosity in typical and dyscalculic development. *Cortex*. https://doi.org/10.1016/j.cortex.2019.07.009

Anobile, G., Burr, D. C., Iaia, M., Marinelli, C. V., Angelelli, P., & Turi, M. (2018). Independent adaptation mechanisms for numerosity and size perception provide evidence against a common sense of magnitude. *Scientific Reports*, *8*(1). https://doi.org/10.1038/s41598-018-31893-6

Anobile, G., Castaldi, E., Moscoso, P. A. M., Burr, D. C., & Arrighi, R. (2020). "Groupitizing": a strategy for numerosity estimation. *Scientific Reports*, *10*(1), 1–9. https://doi.org/10.1038/s41598-020-68111-1

Anobile, G., Castaldi, E., Turi, M., Tinelli, F., & Burr, D. C. (2016). Numerosity but not texture-density discrimination correlates with math ability in children. *Developmental Psychology*. https://doi.org/10.1037/dev0000155

Anobile, G., Cicchini, G. M., & Burr, D. C. (2014). Separate Mechanisms for Perception of Numerosity and Density. *Psychological Science*.

https://doi.org/10.1177/0956797613501520

Anobile, G., Cicchini, G. M., & Burr, D. C. (2016). Number As a Primary Perceptual Attribute: A Review. *Perception*, *45*(1–2), 5–31. https://doi.org/10.1177/0301006615602599

Anobile, G., Cicchini, G. M., Gasperini, F., & Burr, D. C. (2018). Typical numerosity adaptation despite selectively impaired number acuity in dyscalculia. *Neuropsychologia*, *120*. https://doi.org/10.1016/j.neuropsychologia.2018.10.006

Anobile, G., Stievano, P., & Burr, D. C. (2013). Visual sustained attention and numerosity sensitivity correlate with math achievement in children. *Journal of Experimental Child Psychology*. https://doi.org/10.1016/j.jecp.2013.06.006

Anobile, G., Tomaiuolo, F., Campana, S., & Cicchini, G. M. (2020). Three-systems for visual numerosity: A single case study. *Neuropsychologia*. https://doi.org/10.1016/j.neuropsychologia.2019.107259

Anobile, G., Turi, M., Cicchini, G. M., & Burr, D. C. (2012). The effects of cross-sensory attentional demand on subitizing and on mapping number onto space. *Vision Research*. https://doi.org/10.1016/j.visres.2012.06.005

Ansari, D., & Dhital, B. (2006). Age-related changes in the activation of the intraparietal sulcus during nonsymbolic magnitude processing: An event-related functional magnetic resonance imaging study. *Journal of Cognitive Neuroscience*, *18*(11). https://doi.org/10.1162/jocn.2006.18.11.1820

Ansari, D., Lyons, I. M., Van Eimeren, L., & Xu, F. (2007). Linking visual attention and number processing in the brain: The role of the temporo-parietal junction in small and large symbolic and nonsymbolic number comparison. *Journal of Cognitive Neuroscience*, *19*(11). https://doi.org/10.1162/jocn.2007.19.11.1845

Apthorp, D., & Bell, J. (2015). Symmetry is less than meets the eye. In *Current Biology* (Vol. 25, Issue 7). https://doi.org/10.1016/j.cub.2015.02.017

Arrighi, R., Togoli, I., & Burr, D. C. (2014). A generalized sense of number. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1797). https://doi.org/10.1098/rspb.2014.1791

Arsalidou, M., & Taylor, M. J. (2011). Is 2+2=4? Meta-analyses of brain areas needed for numbers and calculations. *NeuroImage*, *54*(3), 2382–2393. https://doi.org/10.1016/j.neuroimage.2010.10.009

Ashcraft, M. H. (2002). Math anxiety: Personal, educational, and cognitive consequences. *Current Directions in Psychological Science*, *11*(5). https://doi.org/10.1111/1467-8721.00196

Ashcraft, M. H., & Faust, M. W. (1994). Mathematics Anxiety And Mental Arithmetic Performance: An Exploratory Investigation. *Cognition and Emotion*, *8*(2). https://doi.org/10.1080/02699939408408931

Ashcraft, M. H., & Kirk, E. P. (2001). The relationships among working memory, math anxiety, and performance. *Journal of Experimental Psychology: General*, *130*(2). https://doi.org/10.1037//0096-3445.130.2.224

Ashcraft, M. H., & Krause, J. A. (2007). Working memory, math performance, and math anxiety. *Psychonomic Bulletin and Review*, *14*(2). https://doi.org/10.3758/BF03194059

Ashcraft, M. H., & Moore, A. M. (2009). Mathematics anxiety and the affective drop in performance. *Journal of Psychoeducational Assessment*, *27*(3). https://doi.org/10.1177/0734282908330580

Ashcraft, M. H., & Ridley, K. S. (2005). Math anxiety and its cognitive consequences: A tutorial review. In *The Handbook of Mathematical Cognition*.

Atkinson, J., Campbell, F. W., & Francis, M. R. (1976). The magic number 4±0: a new look at visual numerosity judgments. *Perception*, *5*, 327–334. https://doi.org/10.1068/p050327

Barlow, H. B., & Földiák, P. (1989). Adaptation and decorrelation in the cortex. In *The Computing Neuron* (Issue 1988).

Barth, H., Kanwisher, N., & Spelke, E. (2003). The construction of large number representations in adults. *Cognition*, *86*(3). https://doi.org/10.1016/S0010-0277(02)00178-6

Beckwith, M., & Restle, F. (1966). Process of enumeration. *Psychological Review*, *73*(5), 437. https://doi.org/10.1037/h0023650

Beilock, S. L., & Maloney, E. A. (2015). Math Anxiety: A Factor in Math Achievement Not to Be Ignored. *Policy Insights from the Behavioral and Brain Sciences*, *2*(1). https://doi.org/10.1177/2372732215601438

Benucci, A., Saleem, A. B., & Carandini, M. (2013). Adaptation maintains population homeostasis in primary visual cortex. *Nature Neuroscience*, *16*(6). https://doi.org/10.1038/nn.3382

Blake, B. J. (1991). *Australian Aboriginal languages: A general introduction.* (2nd ed.). St Lucia: University of Queensland Press.

Bogale, B. A., Kamata, N., Mioko, K., & Sugita, S. (2011). Quantity discrimination in jungle crows, Corvus macrorhynchos. *Animal Behaviour*, *82*(4). https://doi.org/10.1016/j.anbehav.2011.05.025

Bonny, J. W., & Lourenco, S. F. (2013). The approximate number system and its relation to early math achievement: Evidence from the preschool years. *Journal of Experimental Child Psychology*, *114*(3). https://doi.org/10.1016/j.jecp.2012.09.015

Boynton, G. M. (2004). Adaptation and attentional selection. In *Nature Neuroscience* (Vol. 7, Issue 1). https://doi.org/10.1038/nn0104-8

Braham, E. J., & Libertus, M. E. (2018). When approximate number acuity predicts math performance: The moderating role of math anxiety. *PLoS ONE*, *13*(5). https://doi.org/10.1371/journal.pone.0195696

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

Brannon, E. M., Libertus, M. E., Meck, W. H., & Woldorff, M. G. (2008). Electrophysiological measures of time processing in infant and adult brains: Weber's law holds. *Journal of Cognitive Neuroscience*, *20*(2). https://doi.org/10.1162/jocn.2008.20016

Bugden, S., Woldorff, M. G., & Brannon, E. M. (2019). Shared and distinct neural circuitry for nonsymbolic and symbolic double-digit addition. *Human Brain Mapping*, *40*(4), 1328–1343. https://doi.org/10.1002/hbm.24452

Bulthé, J., De Smedt, B., & Op de Beeck, H. P. (2014). Format-dependent representations of symbolic and non-symbolic numbers in the human cortex as revealed by multi-voxel pattern analyses. *NeuroImage*, *87*. https://doi.org/10.1016/j.neuroimage.2013.10.049

Burr, D. C., Anobile, G., & Arrighi, R. (2018). Psychophysical evidence for the number sense. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1740). https://doi.org/10.1098/rstb.2017.0045

Burr, D. C., Anobile, G., & Turi, M. (2011). Adaptation affects both high and low (Subitized) numbers under conditions of high attentional load. *Seeing and Perceiving*. https://doi.org/10.1163/187847511X570097

Burr, D. C., Turi, M., & Anobile, G. (2010). Subitizing but not estimation of numerosity requires attentional resources. *Journal of Vision*. https://doi.org/10.1167/10.6.20

Burr, D., Rocca, E. Della, & Morrone, M. C. (2013). Contextual effects in interval-duration judgements in vision, audition and touch. *Experimental Brain Research*, *230*(1). https://doi.org/10.1007/s00221-013-3632-z

Burr, D., & Ross, J. (2008). A Visual Sense of Number. *Current Biology*, *18*(6), 425–428. https://doi.org/10.1016/j.cub.2008.02.052

Butterworth, B. (1999). *The Mathematical Brain*. Mcmillan.

Butterworth, B. (2008). Numerosity Perception: How Many Speckles on the Hen? *Current Biology*, *18*(9), R388–R389. https://doi.org/10.1016/j.cub.2008.03.014

Butterworth, B. (2019). *Dyscalculia: From science to education*. Routledge.

Butterworth, B., Varma, S., & Laurillard, D. (2011). Dyscalculia: From brain to education. In *Science* (Vol. 332, Issue 6033). https://doi.org/10.1126/science.1201536

Cabeza, R., Ciaramelli, E., & Moscovitch, M. (2012). Cognitive contributions of the ventral parietal cortex: An integrative theoretical account. *Trends in Cognitive Sciences*, *16*(6), 338–352. https://doi.org/10.1016/j.tics.2012.04.008

Camos, V., & Tillmann, B. (2008). Discontinuity in the enumeration of sequentially presented auditory and visual stimuli. *Cognition*. https://doi.org/10.1016/j.cognition.2007.11.002

Cantlon, J. F., & Brannon, E. M. (2006). Shared system for ordering small and large numbers in monkeys and humans. *Psychological Science*, *17*(5). https://doi.org/10.1111/j.1467-9280.2006.01719.x

Cantlon, J. F., Brannon, E. M., Carter, E. J., & Pelphrey, K. A. (2006). Functional imaging of numerical processing in adults and 4-y-old children. *PLoS Biology*, *4*(5). https://doi.org/10.1371/journal.pbio.0040125

Cantlon, J. F., Platt, M. L., & Brannon, E. M. (2009). Beyond the number domain. In *Trends in Cognitive Sciences* (Vol. 13, Issue 2). https://doi.org/10.1016/j.tics.2008.11.007

Carey, E., Hill, F., Devine, A., & Szücs, D. (2016). The chicken or the egg? The direction of the relationship between mathematics anxiety and mathematics performance. In *Frontiers in Psychology* (Vol. 6, Issue JAN). https://doi.org/10.3389/fpsyg.2015.01987

Castaldi, E., Aagten-Murphy, D., Tosetti, M., Burr, D. C., & Morrone, M. C. (2016). Effects of adaptation on numerosity decoding in the human brain. *NeuroImage*, *143*, 364–377. https://doi.org/10.1016/j.neuroimage.2016.09.020

Castaldi, E., Mirassou, A., Dehaene, S., Piazza, M., & Eger, E. (2018). Asymmetrical interference between number and item size perception provides evidence for a domain specific impairment in dyscalculia. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0209256

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. https://doi.org/10.7554/eLife.45160

Castaldi, E., Turi, M., Gassama, S., Piazza, M., & Eger, E. (2020). Excessive visual crowding effects in developmental dyscalculia. *Journal of Vision*. https://doi.org/10.1167/jov.20.8.7

Castaldi, E., Vignaud, A., & Eger, E. (2020). Mapping subcomponents of numerical cognition in relation to functional and anatomical landmarks of human parietal cortex. *NeuroImage*, *221*, 117210. https://doi.org/10.1016/j.neuroimage.2020.117210

Castronovo, J., & Göbel, S. M. (2012). Impact of high mathematics education on the number sense. *PloS One*, *7*(4). https://doi.org/10.1371/journal.pone.0033832

Chang, H., & Beilock, S. L. (2016). The math anxiety-math performance link and its relation to individual and environmental factors: A review of current behavioral and psychophysiological research. In *Current Opinion in Behavioral Sciences* (Vol. 10). https://doi.org/10.1016/j.cobeha.2016.04.011

Chochon, F., Cohen, L., Moortele, P. V. D., & Dehaene, S. (1999). Differential contributions of the left and right inferior parietal lobules to number processing. *Journal of Cognitive Neuroscience*, *11*(6), 617–630.

Cicchini, G. M., Anobile, G., & Burr, D. C. (2014). Compressive mapping of number to space reflects dynamic encoding mechanisms, not static logarithmic transform. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.1402785111

Cicchini, G. M., Anobile, G., & Burr, D. C. (2016). Spontaneous perception of numerosity in humans. *Nature Communications*, *7*. https://doi.org/10.1038/ncomms12536

Cicchini, G. M., Anobile, G., & Burr, D. C. (2019). Spontaneous representation of numerosity in typical and dyscalculic development. *Cortex*, *114*. https://doi.org/10.1016/j.cortex.2018.11.019

Cicchini, G. M., Arrighi, R., Cecchetti, L., Giusti, M., & Burr, D. C. (2012). Optimal encoding of interval timing in expert percussionists. *Journal of Neuroscience*. https://doi.org/10.1523/JNEUROSCI.3411-11.2012

Cicchini, G. M., & Burr, D. C. (2018). Serial effects are optimal. In *The Behavioral and brain sciences*. https://doi.org/10.1017/S0140525X18001395

Cicchini, G. M., Mikellidou, K., & Burr, D. (2017). Serial dependencies act directly on perception. *Journal of Vision*, *17*(14), 6. https://doi.org/10.1167/17.14.6

Ciccione, L., & Dehaene, S. (2020). Grouping mechanisms in numerosity perception. *Open Mind*, *4*, 102–118. https://doi.org/10.31234/osf.io/p6ryv

Cipolotti, L., Butterworth, B., & Denes, G. (1991). A specific deficit for numbers in a case of dense acalculia. *Brain*, *114*(6), 2619–2637. https://doi.org/10.1093/brain/114.6.2619

Clifford, C. W. G., & Rhodes, G. (2005). Fitting the Mind to the World: Adaptation and After-Effects in High-Level Vision. In *Fitting the Mind to* *the World: Adaptation and After-Effects in High-Level Vision*. https://doi.org/10.1093/acprof:oso/9780198529699.001.0001

Cohen Kadosh, R., Bahrami, B., Walsh, V., Butterworth, B., Popescu, T., & Price, C. J. (2011). Specialization in the human brain: The case of numbers. *Frontiers in Human Neuroscience*, *JULY*. https://doi.org/10.3389/fnhum.2011.00062

Colomé, À. (2019). Representation of numerical magnitude in math-anxious individuals. *Quarterly Journal of Experimental Psychology*, *72*(3). https://doi.org/10.1177/1747021817752094

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

Cordes, S., Gelman, R., Gallistel, C. R., & Whalen, J. (2001). Variability signatures distinguish verbal from nonverbal counting for both large and small numbers. *Psychonomic Bulletin and Review*. https://doi.org/10.3758/BF03196206

Coubart, A., Izard, V., Spelke, E. S., Marie, J., & Streri, A. (2014). Dissociation between small and large numerosities in newborn infants. *Developmental Science*, *17*(1). https://doi.org/10.1111/desc.12108

Dakin, S. C., Tibber, M. S., Greenwood, J. A., Kingdom, F. A. A., & Morgan, M. J. (2011). A common visual metric for approximate number and density. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(49). https://doi.org/10.1073/pnas.1113195108

Damarla, S. R., & Just, M. A. (2013). Decoding the representation of numerical values from brain activation patterns. *Human Brain Mapping*, *34*(10). https://doi.org/10.1002/hbm.22087

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. In *Trends in Neuroscience and Education* (Vol. 2, Issue 2). https://doi.org/10.1016/j.tine.2013.06.001

De Smedt, B., Verschaffel, L., & Ghesquière, P. (2009). The predictive value of numerical magnitude comparison for individual differences in mathematics achievement. *Journal of Experimental Child Psychology*, *103*(4). https://doi.org/10.1016/j.jecp.2009.01.010

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition*. https://doi.org/10.1016/0010-0277(92)90049-N

Dehaene, S. (2003). The neural basis of the Weber-Fechner law: A logarithmic mental number line. In *Trends in Cognitive Sciences* (Vol. 7, Issue 4). https://doi.org/10.1016/S1364-6613(03)00055-X

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

Dehaene, S., & Cohen, L. (1997). Cerebral pathways for calculation: Double dissociation between rote verbal and quantitative knowledge of arithmetic. *Cortex*, *33*(2). https://doi.org/10.1016/S0010-9452(08)70002-9

Dehaene, S., Dehaene-Lambertz, G., & Cohen, L. (1998). Abstract representations of numbers in the animal and human brain. *Trends in Neurosciences*, *21*(8). https://doi.org/10.1016/S0166-2236(98)01263-6

Dehaene, S., Piazza, M., Pinel, P., & Cohen, L. (2003). Three parietal circuits for number processing. *Cognitive Neuropsychology*, *20*(3–6), 487–506. https://doi.org/10.1080/02643290244000239

Dehaene, S., Spelke, E., Pinel, P., Stanescu, R., & Tsivkin, S. (1999). Sources of mathematical thinking: Behavioral and brain-imaging evidence. *Science*, *284*(5416), 970–974. https://doi.org/10.1126/science.284.5416.970

Delazer, M., & Benke, T. (1997). Arithmetic facts without meaning. *Cortex*, *33*(4). https://doi.org/10.1016/S0010-9452(08)70727-5

Delazer, M., Domahs, F., Bartha, L., Brenneis, C., Lochy, A., Trieb, T., & Benke, T. (2003). Learning complex arithmetic - An fMRI study. *Cognitive Brain Research*, *18*(1), 76–88. https://doi.org/10.1016/j.cogbrainres.2003.09.005

Demeyere, N., Rotshtein, P., & Humphreys, G. W. (2014). Common and dissociated mechanisms for estimating large and small dot arrays: Value-specific fMRI adaptation. *Human Brain Mapping*, *35*(8). https://doi.org/10.1002/hbm.22453

Destrieux, C., Fischl, B., Dale, A., & Halgren, E. (2010). Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *NeuroImage*, *53*(1), 1–15. https://doi.org/10.1016/j.neuroimage.2010.06.010

Dietrich, J. F., Huber, S., Moeller, K., & Klein, E. (2015). The influence of math anxiety on symbolic and non-symbolic magnitude processing. *Frontiers in Psychology*, *6*(OCT). https://doi.org/10.3389/fpsyg.2015.01621

Dormal, V., Andres, M., Dormal, G., & Pesenti, M. (2010). Mode-dependent and mode-independent representations of numerosity in the right intraparietal sulcus. *NeuroImage*, *52*(4), 1677–1686. https://doi.org/10.1016/j.neuroimage.2010.04.254

Dormal, V., Dormal, G., Joassin, F., & Pesenti, M. (2012). A common right fronto-parietal network for numerosity and duration processing: An fMRI study. *Human Brain Mapping*, *33*(6), 1490–1501.

Dormal, V., & Pesenti, M. (2009). Common and specific contributions of the intraparietal sulci to numerosity and length processing. *Human Brain Mapping*, *30*(8). https://doi.org/10.1002/hbm.20677

Duffau, H., Denvil, D., Lopes, M., Gasparini, F., Cohen, L., Capelle, L., & Van Effenterre, R. (2002). Intraoperative mapping of the cortical areas involved in multiplication and subtraction: An electrostimulation study in a patient with a left parietal glioma. *Journal of Neurology Neurosurgery and Psychiatry*, *73*(6). https://doi.org/10.1136/jnnp.73.6.733

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., Pagani, L. S., Feinstein, L., Engel, M., Brooks-Gunn, J., Sexton, H., Duckworth, K., & Japel, C. (2007). School Readiness and Later Achievement. *Developmental Psychology*, *43*(6). https://doi.org/10.1037/0012-1649.43.6.1428

Durgin, F. H. (1995). Texture Density Adaptation and the Perceived Numerosity and Distribution of Texture. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(1). https://doi.org/10.1037/0096-1523.21.1.149

Durgin, F. H. (2008). Texture density adaptation and visual number revisited. In *Current Biology* (Vol. 18, Issue 18). https://doi.org/10.1016/j.cub.2008.07.053

Eger, E., Michel, V., Thirion, B., Amadon, A., Dehaene, S., & Kleinschmidt, A. (2009). Deciphering Cortical Number Coding from Human Brain Activity Patterns. *Current Biology*, *19*(19). https://doi.org/10.1016/j.cub.2009.08.047

Eger, E., Pinel, P., Dehaene, S., & Kleinschmidt, A. (2015). Spatially invariant coding of numerical information in functionally defined subregions of human parietal cortex. *Cerebral Cortex*, *25*(5), 1319–1329. https://doi.org/10.1093/cercor/bht323

Eger, E., Sterzer, P., Russ, M. O., Giraud, A. L., & Kleinschmidt, A. (2003). A supramodal number representation in human intraparietal cortex. *Neuron*, *37*(4). https://doi.org/10.1016/S0896-6273(03)00036-9

Egeth, H. E., Leonard, C. J., & Palomares, M. (2008). The role of attention in subitizing: Is the magical number 1? *Visual Cognition*, *16*(4), 463–473. https://doi.org/10.1080/13506280801937939

Embretson, S. E., & Reise, S. P. (2000). *Item Response Theory for Psychologists*. Lawrence Erlbaum Associates.

Evans, J. (2002). Adults' Mathematical Thinking and Emotions. In *Adults' Mathematical Thinking and Emotions*. https://doi.org/10.4324/9780203185896

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*(1). https://doi.org/10.1016/j.jecp.2014.01.013

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. In *Trends in Cognitive Sciences* (Vol. 8, Issue 7). https://doi.org/10.1016/j.tics.2004.05.002

Feigenson, L., Libertus, M. E., & Halberda, J. (2013). Links Between the Intuitive Sense of Number and Formal Mathematics Ability. *Child Development Perspectives*, *7*(2). https://doi.org/10.1111/cdep.12019

Ferguson, A. M., Maloney, E. A., Fugelsang, J., & Risko, E. F. (2015). On the relation between math and spatial ability: The case of math anxiety. *Learning and Individual Differences*, *39*. https://doi.org/10.1016/j.lindif.2015.02.007

Firestone, C., & Scholl, B. J. (2016). Cognition does not affect perception: Evaluating the evidence for "top-down" effects. *Behavioral and Brain Sciences*, *39*, e229. https://doi.org/10.1017/S0140525X15000965

Fischer, J., & Whitney, D. (2014). Serial dependence in visual perception. *Nature Neuroscience*, *17*(5), 738–743. https://doi.org/10.1038/nn.3689

Fischl, B., Van Der Kouwe, A., Destrieux, C., Halgren, E., Ségonne, F., Salat, D. H., Busa, E., Seidman, L. J., Goldstein, J., Kennedy, D., Caviness, V., Makris, N., Rosen, B., & Dale, A. M. (2004). Automatically Parcellating the Human Cerebral Cortex. *Cerebral Cortex*, *14*(1), 11–22. https://doi.org/10.1093/cercor/bhg087

Fornaciai, M., & Park, J. (2019). Serial dependence generalizes across different stimulus formats, but not different sensory modalities. *Vision Research*, *160*, 108–115. https://doi.org/10.1016/j.visres.2019.04.011

Frisby, J. P. (1979). *Seeing: Illusion, Brain and Mind.* Oxford University Press.

Gallagher, R. M., Suddendorf, T., & Arnold, D. H. (2019). Confidence as a diagnostic tool for perceptual aftereffects. *Scientific Reports*, *9*(1). https://doi.org/10.1038/s41598-019-43170-1

Galli, S., Chiesi, F., & Primi, C. (2011). Measuring mathematical ability needed for “non-mathematical” majors: The construction of a scale applying IRT and differential item functioning across educational contexts. *Learning and Individual Differences*, *21*(4). https://doi.org/10.1016/j.lindif.2011.04.005

Gallistel, C. R., & Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition*, *44*(1–2). https://doi.org/10.1016/0010-0277(92)90050-R

Gallistel, C. R., & Gelman, R. (2000). Non-verbal numerical cognition: From reals to integers. In *Trends in Cognitive Sciences* (Vol. 4, Issue 2). https://doi.org/10.1016/S1364-6613(99)01424-2

Gerstmann, J. (1940). Syndrome of finger agnosia, disorientation for right and left, agraphia and acalculia: Local diagnostic value. *Archives of Neurology And Psychiatry*, *44*(2), 398–408. https://doi.org/10.1001/archneurpsyc.1940.02280080158009

Gilmore, C., McCarthy, S. E., & Spelke, E. S. (2010). Non-symbolic arithmetic abilities and mathematics achievement in the first year of formal schooling. *Cognition*, *115*(3). https://doi.org/10.1016/j.cognition.2010.02.002

Göbel, S., Walsh, V., & Rushworth, M. F. S. (2001). The mental number line and the human angular gyrus. *NeuroImage*, *14*(6), 1278–1289. https://doi.org/10.1006/nimg.2001.0927

Gómez-Velázquez, F. R., Berumen, G., & González-Garrido, A. A. (2015). Comparisons of numerical magnitudes in children with different levels of mathematical achievement. An ERP study. *Brain Research*, *1627*. https://doi.org/10.1016/j.brainres.2015.09.009

Gordon, P. (2004). Numerical cognition without words: Evidence from Amazonia. *Science*, *306*(5695). https://doi.org/10.1126/science.1094492

Grabner, R. H., Ansari, D., Koschutnig, K., Reishofer, G., & Ebner, F. (2013). The function of the left angular gyrus in mental arithmetic: Evidence from the associative confusion effect. *Human Brain Mapping*, *34*(5), 1013–1024. https://doi.org/10.1002/hbm.21489

Grabner, R. H., Ansari, D., Koschutnig, K., Reishofer, G., Ebner, F., & Neuper, C. (2009). To retrieve or to calculate? Left angular gyrus mediates the retrieval of arithmetic facts during problem solving. *Neuropsychologia*, *47*(2), 604–608. https://doi.org/10.1016/j.neuropsychologia.2008.10.013

Grabner, R. H., Ansari, D., Reishofer, G., Stern, E., Ebner, F., & Neuper, C. (2007). Individual differences in mathematical competence predict parietal brain activation during mental calculation. *NeuroImage*, *38*(2), 346–356. https://doi.org/10.1016/j.neuroimage.2007.07.041

Grabner, R. H., Ischebeck, A., Reishofer, G., Koschutnig, K., Delazer, M., Ebner, F., & Neuper, C. (2009). Fact learning in complex arithmetic and figural-spatial tasks: The role of the angular gyrus and its relation to mathematical competence. *Human Brain Mapping*, *30*(9), 2936–2952. https://doi.org/10.1002/hbm.20720

Gray, S. A., & Reeve, R. A. (2014). Preschoolers’ dot enumeration abilities are markers of their arithmetic competence. *PLoS ONE*, *9*(4). https://doi.org/10.1371/journal.pone.0094428

Hagler, D. J., Saygin, A. P., & Sereno, M. I. (2006). Smoothing and cluster thresholding for cortical surface-based group analysis of fMRI data.

NeuroImage, *33*(4), 1093–1103. https://doi.org/10.1016/j.neuroimage.2006.07.036

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). https://doi.org/10.1037/a0012682

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 of the United States of America*, *109*(28). https://doi.org/10.1073/pnas.1200196109

Halberda, J., Mazzocco, M. M. M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature*, *455*(7213). https://doi.org/10.1038/nature07246

Hart, S. A., Logan, J. A. R., Thompson, L., Kovas, Y., McLoughlin, G., & Petrill, S. A. (2016). A latent profile analysis of math achievement, numerosity, and math anxiety in twins. *Journal of Educational Psychology*, *108*(2). https://doi.org/10.1037/edu0000045

Harvey, B. M., Klein, B. P., Petridou, N., & Dumoulin, S. O. (2013). Topographic representation of numerosity in the human parietal cortex. *Science*, *341*(6150). https://doi.org/10.1126/science.1239052

Hayes, A. F. (2009). Beyond Baron and Kenny: Statistical mediation analysis in the new millennium. *Communication Monographs*, *76*(4). https://doi.org/10.1080/03637750903310360

Hayes, A. F. (2013). *INDIRECT Macro for SPSS.*

Hayes, A. F., & Scharkow, M. (2013). The Relative Trustworthiness of Inferential Tests of the Indirect Effect in Statistical Mediation Analysis: Does Method Really Matter? *Psychological Science*, *24*(10). https://doi.org/10.1177/0956797613480187

He, L., Zhou, K., Zhou, T., He, S., & Chen, L. (2015). Topology-defined units in numerosity perception. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(41). https://doi.org/10.1073/pnas.1512408112

Hembree, R. (1990). The Nature, Effects, and Relief of Mathematics Anxiety. *Journal for Research in Mathematics Education*, *21*(1). https://doi.org/10.2307/749455

Henik, A., Gliksman, Y., Kallai, A., & Leibovich, T. (2017). Size Perception and the Foundation of Numerical Processing. *Current Directions in Psychological Science*, *26*(1). https://doi.org/10.1177/0963721416671323

Hollingworth, H. L. (1910). The Central Tendency of Judgment. *The Journal of Philosophy, Psychology and Scientific Methods*. https://doi.org/10.2307/2012819

Holloway, I. D., & Ansari, D. (2009). Mapping numerical magnitudes onto symbols: The numerical distance effect and individual differences in children's mathematics achievement. *Journal of Experimental Child Psychology*, *103*(1). https://doi.org/10.1016/j.jecp.2008.04.001

Hopko, D. R., Mahadevan, R., Bare, R. L., & Hunt, M. K. (2003). The Abbreviated Math Anxiety Scale (AMAS): Construction, validity, and reliability. *Assessment*, *10*(2).

https://doi.org/10.1177/1073191103010002008

Huntley-Fenner, G. (2001). Children's understanding of number is similar to adults' and rats': Numerical estimation by 5-7-year-olds. *Cognition*, *78*(3). https://doi.org/10.1016/S0010-0277(00)00122-0

Hutchison, J. E., Ansari, D., Zheng, S., De Jesus, S., & Lyons, I. M. (2020). The relation between subitizable symbolic and non-symbolic number processing over the kindergarten school year. *Developmental Science*, *23*(2). https://doi.org/10.1111/desc.12884

Hyde, D. C., Khanum, S., & Spelke, E. S. (2014). Brief non-symbolic, approximate number practice enhances subsequent exact symbolic arithmetic in children. *Cognition*, *131*(1). https://doi.org/10.1016/j.cognition.2013.12.007

Hyde, D. C., & Spelke, E. S. (2011). Neural signatures of number processing in human infants: Evidence for two core systems underlying numerical cognition. *Developmental Science*, *14*(2). https://doi.org/10.1111/j.1467-7687.2010.00987.x

Inglis, M., Attridge, N., Batchelor, S., & Gilmore, C. (2011). Non-verbal number acuity correlates with symbolic mathematics achievement: But only in children. *Psychonomic Bulletin and Review*, *18*(6). https://doi.org/10.3758/s13423-011-0154-1

Ischebeck, A., Zamarian, L., Egger, K., Schocke, M., & Delazer, M. (2007). Imaging early practice effects in arithmetic. *NeuroImage*, *36*(3), 993–1003. https://doi.org/10.1016/j.neuroimage.2007.03.051

Ischebeck, A., Zamarian, L., Schocke, M., & Delazer, M. (2009). Flexible transfer of knowledge in mental arithmetic - An fMRI study. *NeuroImage*, *44*(3), 1103–1112. https://doi.org/10.1016/j.neuroimage.2008.10.025

Iuculano, T., Tang, J., Hall, C. W. B., & Butterworth, B. (2008). Core information processing deficits in developmental dyscalculia and low numeracy. *Developmental Science*, *11*(5). https://doi.org/10.1111/j.1467-7687.2008.00716.x

Izard, V., Dehaene-Lambertz, G., & Dehaene, S. (2008). Distinct cerebral pathways for object identity and number in human infants. *PLoS Biology*, *6*(2). https://doi.org/10.1371/journal.pbio.0060011

Jackson, M., & Warrington, E. K. (1986). Arithmetic Skills in Patients with Unilateral Cerebral Lesions. *Cortex*, *22*(4), 611–620. https://doi.org/10.1016/S0010-9452(86)80020-X

Jacob, S. N., & Nieder, A. (2009). Tuning to non-symbolic proportions in the human frontoparietal cortex. *European Journal of Neuroscience*, *30*(7). https://doi.org/10.1111/j.1460-9568.2009.06932.x

Jazayeri, M., & Shadlen, M. N. (2010). Temporal context calibrates interval timing. *Nature Neuroscience*. https://doi.org/10.1038/nn.2590

Jeffreys, H. (1998). *Theory of probability* (3rd ed.). Clarendon Press.

Jevons, W. S. (1871). The power of numerical discrimination. *Nature*, *3*, 281–282.

Jones, M. R., & Mcauley, J. D. (2005). Time judgments in global temporal contexts. *Perception and Psychophysics*, *67*(3). https://doi.org/10.3758/BF03193320

Jordan, K. E., & Brannon, E. M. (2006). The multisensory representation of number in infancy. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(9). https://doi.org/10.1073/pnas.0508107103

Kaufman, E. L., Lord, M. W., Reese, T. W., & Volkmann, J. (1949). The discrimination of visual number. *The American Journal of Psychology*, *62*(4), 498–525.

Klein, A., & Starkey, P. (1988). Universals in the development of early arithmetic cognition. *New Directions for Child and Adolescent Development*. https://doi.org/10.1002/cd.23219884103

Kohn, A. (2007). Visual adaptation: Physiology, mechanisms, and functional benefits. In *Journal of Neurophysiology* (Vol. 97, Issue 5). https://doi.org/10.1152/jn.00086.2007

Konen, C. S., & Kastner, S. (2008). Representation of eye movements and stimulus motion in topographically organized areas of human posterior parietal cortex. *Journal of Neuroscience*, *28*(33). https://doi.org/10.1523/JNEUROSCI.1930-08.2008

Kramer, P., Di Bono, M. G., & Zorzi, M. (2011). Numerosity estimation in visual stimuli in the absence of luminance-based cues. *PLoS ONE*, *6*(2). https://doi.org/10.1371/journal.pone.0017378

Krueger, L. E. (1984). Perceived numerosity: A comparison of magnitude production, magnitude estimation, and discrimination judgments. *Perception & Psychophysics*, *35*(6). https://doi.org/10.3758/BF03205949

Landerl, K., Bevan, A., & Butterworth, B. (2004). Developmental dyscalculia and basic numerical capacities: A study of 8-9-year-old students. *Cognition*, *93*(2). https://doi.org/10.1016/j.cognition.2003.11.004

Laplace, P. S. (1812). Théorie analytique des Probabilités. In *2*. Courcier Imprimeur.

Lappin, J. S., Shelton, A. L., & Rieser, J. J. (2006). Environmental context influences visually perceived distance. *Perception and Psychophysics*, *68*(4). https://doi.org/10.3758/BF03208759

Lasne, G., Piazza, M., Dehaene, S., Kleinschmidt, A., & Eger, E. (2019). Discriminability of numerosity-evoked fMRI activity patterns in human intra-parietal cortex reflects behavioral numerical acuity. *Cortex*, *114*, 90–101. https://doi.org/10.1016/j.cortex.2018.03.008

Lee, K. M. (2000). Cortical areas differentially involved in multiplication and subtraction: A functional magnetic resonance imaging study and correlation with a case of selective acalculia. *Annals of Neurology*, *48*(4), 657–661. https://doi.org/10.1002/1531-8249(200010)48:4<657::AID-ANA13>3.0.CO;2-K

Leibovich, T., & Ansari, D. (2016). The Symbol-Grounding Problem in Numerical Cognition: A Review of Theory, Evidence, and Outstanding Questions. *Canadian Journal of Experimental Psychology*, *70*(1). https://doi.org/10.1037/cep0000070

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*. https://doi.org/10.1017/S0140525X16000960

Leopold, D. A., Rhodes, G., Müller, K. M., & Jeffery, L. (2005). The dynamics of visual adaptation to faces. *Proceedings of the Royal Society B: Biological*

Sciences, *272*(1566), 897–904. https://doi.org/10.1098/rspb.2004.3022

Liberman, A., Fischer, J., & Whitney, D. (2014). Serial dependence in the perception of faces. *Current Biology*. https://doi.org/10.1016/j.cub.2014.09.025

Libertus, M. E., Feigenson, L., & Halberda, J. (2011). Preschool acuity of the approximate number system correlates with school math ability. *Developmental Science*, *14*(6). https://doi.org/10.1111/j.1467-7687.2011.01080.x

Libertus, M. E., Feigenson, L., & Halberda, J. (2013). Is approximate number precision a stable predictor of math ability? *Learning and Individual Differences*, *25*. https://doi.org/10.1016/j.lindif.2013.02.001

Libertus, M. E., Odic, D., & Halberda, J. (2012). Intuitive sense of number correlates with math scores on college-entrance examination. *Acta Psychologica*, *141*(3). https://doi.org/10.1016/j.actpsy.2012.09.009

Lindskog, M., & Winman, A. (2016). No evidence of learning in non-symbolic numerical tasks - A comment on Park and Brannon (2014). In *Cognition*. https://doi.org/10.1016/j.cognition.2016.01.005

Lindskog, M., Winman, A., & Poom, L. (2017). Individual differences in nonverbal number skills predict math anxiety. *Cognition*, *159*. https://doi.org/10.1016/j.cognition.2016.11.014

Lipton, J. S., & Spelke, E. S. (2003). Origins of Number Sense: Large-number discrimination in Human Infants. *Psychological Science*, *14*(5). https://doi.org/10.1111/1467-9280.01453

Lonnemann, J., Linkersdörfer, J., Hasselhorn, M., & Lindberg, S. (2011). Symbolic and non-symbolic distance effects in children and their connection with arithmetic skills. *Journal of Neurolinguistics*, *24*(5). https://doi.org/10.1016/j.jneuroling.2011.02.004

Lourenco, S. F., Bonny, J. W., Fernandez, E. P., & Rao, S. (2012). Nonsymbolic number and cumulative area representations contribute shared and unique variance to symbolic math competence. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(46). https://doi.org/10.1073/pnas.1207212109

Lyons, I. M., & Beilock, S. L. (2011). Numerical ordering ability mediates the relation between number-sense and arithmetic competence. *Cognition*. https://doi.org/10.1016/j.cognition.2011.07.009

Lyons, I. M., Price, G. R., Vaessen, A., Blomert, L., & Ansari, D. (2014). Numerical predictors of arithmetic success in grades 1-6. *Developmental Science*, *17*(5). https://doi.org/10.1111/desc.12152

Ma, X. (1999). A meta-analysis of the relationship between anxiety toward mathematics and achievement in mathematics. *Journal for Research in Mathematics Education*, *30*(5). https://doi.org/10.2307/749772

Ma, X., & Kishor, N. (1997). Assessing the relationship between attitude toward mathematics and achievement in mathematics: A meta-analysis. *Journal for Research in Mathematics Education*, *28*(1). https://doi.org/10.2307/749662

Ma, X., & Xu, J. (2004). The causal ordering of mathematics anxiety and mathematics achievement: A longitudinal panel analysis. *Journal of Adolescence*, *27*(2). https://doi.org/10.1016/j.adolescence.2003.11.003

MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V.

(2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods*, *7*(1). https://doi.org/10.1037/1082-989X.7.1.83

MacKinnon, D. P., Lockwood, C. M., & Williams, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research*, *39*(1). https://doi.org/10.1207/s15327906mbr3901_4

Malanchini, M., Rimfeld, K., Shakeshaft, N. G., Rodic, M., Schofield, K., Selzam, S., Dale, P. S., Petrill, S. A., & Kovas, Y. (2017). The genetic and environmental aetiology of spatial, mathematics and general anxiety. *Scientific Reports*, *7*. https://doi.org/10.1038/srep42218

Maldonado Moscoso, P. A., Castaldi, E., Burr, D. C., Arrighi, R., & Anobile, G. (2020). Grouping strategies in number estimation extend the subitizing range. *Scientific Reports*, *10*(1), 1–10. https://doi.org/10.1038/s41598-020-71871-5

Maloney, E. A. (2016). Math anxiety: Causes, consequences, and remediation. In *Handbook of Motivation at School: Second Edition*. https://doi.org/10.4324/9781315773384

Maloney, E. A., Ansari, D., & Fugelsang, J. A. (2011). The effect of mathematics anxiety on the processing of numerical magnitude. *Quarterly Journal of Experimental Psychology*, *64*(1). https://doi.org/10.1080/17470218.2010.533278

Maloney, E. A., Risko, E. F., Ansari, D., & Fugelsang, J. (2010). Mathematics anxiety affects counting but not subitizing during visual enumeration. *Cognition*, *114*(2). https://doi.org/10.1016/j.cognition.2009.09.013

Mamassian, P. (2016). Visual Confidence. *Annual Review of Vision Science*, *2*, 459–481. https://doi.org/10.1146/annurev-vision-111815-114630

Mandler, G., & Shebo, B. J. (1982). Subitizing: An analysis of its component processes. *Journal of Experimental Psychology: General*. https://doi.org/10.1037/0096-3445.111.1.1

Maniscalco, B., & Lau, H. (2014). Signal detection theory analysis of type 1 and type 2 data: Meta-d′, response-specific meta-d′, and the unequal variance SDT model. In *The Cognitive Neuroscience of Metacognition* (pp. 25–66). https://doi.org/10.1007/978-3-642-45190-4_3

Mazzocco, M. M. M., Feigenson, L., & Halberda, J. (2011). Impaired acuity of the approximate number system underlies mathematical learning disability (dyscalculia). *Child Development*, *82*(4). https://doi.org/10.1111/j.1467-8624.2011.01608.x

McAuley, J. D., & Jones, M. R. (2003). Modeling Effects of Rhythmic Context on Perceived Duration: A Comparison of Interval and Entrainment Approaches to Short-Interval Timing. *Journal of Experimental Psychology: Human Perception and Performance*, *29*(6). https://doi.org/10.1037/0096-1523.29.6.1102

McComb, K., Packer, C., & Pusey, A. (1994). Roaring and numerical assessment in contests between groups of female lions, Panthera leo. *Animal Behaviour*, *47*(2). https://doi.org/10.1006/anbe.1994.1052

Meck, W. H., & Church, R. M. (1983). A mode control model of counting and timing processes. *Journal of Experimental Psychology: Animal Behavior*

Processes, *9*(3). https://doi.org/10.1037/0097-7403.9.3.320

Menon, V., Rivera, S. M., White, C. D., Glover, G. H., & Reiss, A. L. (2000). Dissociating prefrontal and parietal cortex activation during arithmetic processing. *NeuroImage*, *12*(4), 357–365. https://doi.org/10.1006/nimg.2000.0613

Mollon, J. D. (1974). After-effects and the brain. *New Scientist*, *61*(886), 479–482.

Morgan, M. J., Dillenburger, B., Raphael, S., & Solomon, J. A. (2011). Observers can voluntarily shift their psychometric functions without losing sensitivity. *Attention, Perception, and Psychophysics*, *74*, 185–193. https://doi.org/10.3758/s13414-011-0222-7

Morgan, M. J., Raphael, S., Tibber, M. S., & Dakin, S. C. (2014). A texture-processing model of the "visual sense of number." *Proceedings of the Royal Society B: Biological Sciences*, *281*(1790). https://doi.org/10.1098/rspb.2014.1137

Morsanyi, K., Busdraghi, C., & Primi, C. (2014). Mathematical anxiety is linked to reduced cognitive reflection: A potential road from discomfort in the mathematics classroom to susceptibility to biases. *Behavioral and Brain Functions*, *10*(1). https://doi.org/10.1186/1744-9081-10-31

Nieder, A. (2005). Counting on neurons: The neurobiology of numerical competence. In *Nature Reviews Neuroscience* (Vol. 6, Issue 3). https://doi.org/10.1038/nrn1626

Nieder, A. (2016). The neuronal code for number. *Nature Reviews Neuroscience*, *17*(6), 366. https://doi.org/10.1038/nrn.2016.40

Nieder, A., & Dehaene, S. (2009). Representation of Number in the Brain. *Annual Review of Neuroscience*, *32*, 185–208. https://doi.org/10.1146/annurev.neuro.051508.135550

Nieder, A., Freedman, D. J., & Miller, E. K. (2002). Representation of the quantity of visual items in the primate prefrontal cortex. *Science*, *297*(5587). https://doi.org/10.1126/science.1072493

Nieder, A., & Miller, E. K. (2004). A parieto-frontal network for visual numerical information in the monkey. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(19). https://doi.org/10.1073/pnas.0402239101

Núñez-Peña, M. I., & Suaŕez-Pellicioni, M. (2014). Less precise representation of numerical magnitude in high math-anxious individuals: An ERP study of the size and distance effects. *Biological Psychology*, *103*. https://doi.org/10.1016/j.biopsycho.2014.09.004

Núñez-Peña, M. I., & Suaŕez-Pellicioni, M. (2015). Processing of multi-digit additions in high math-anxious individuals: psychophysiological evidence. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01268

Obersteiner, A., Reiss, K., & Ufer, S. (2013). How training on exact or approximate mental representations of number can enhance first-grade students' basic number processing and arithmetic skills. *Learning and Instruction*, *23*(1). https://doi.org/10.1016/j.learninstruc.2012.08.004

Odic, D., Libertus, M. E., Feigenson, L., & Halberda, J. (2013). Developmental change in the acuity of approximate number and area representations. *Developmental Psychology*, *49*(6). https://doi.org/10.1037/a0029472

Olivers, C. N. L., & Watson, D. G. (2008). Subitizing requires attention. *Visual Cognition*. https://doi.org/10.1080/13506280701825861

Park, J., Bermudez, V., Roberts, R. C., & Brannon, E. M. (2016). Non-symbolic approximate arithmetic training improves math performance in preschoolers. *Journal of Experimental Child Psychology*, *152*. https://doi.org/10.1016/j.jecp.2016.07.011

Park, J., & Brannon, E. M. (2013). Training the Approximate Number System Improves Math Proficiency. *Psychological Science*, *24*(10). https://doi.org/10.1177/0956797613482944

Park, J., & Brannon, E. M. (2014). Improving arithmetic performance with number sense training: An investigation of underlying mechanism. *Cognition*, *133*(1). https://doi.org/10.1016/j.cognition.2014.06.011

Parsons, S., & Bynner, J. (2005). Does Numeracy Matter More ? In *National Research and Development Centre for Adult Literacy and Numeracy*. https://doi.org/1905188090

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Pelli, D. G. (1985). Uncertainty explains many aspects of visual contrast detection and discrimination. *Journal of the Optical Society of America A*, *2*(9), 1508–1532. https://doi.org/10.1364/josaa.2.001508

Piazza, M. (2010). Neurocognitive start-up tools for symbolic number representations. In *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2010.09.008

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). https://doi.org/10.1016/j.cognition.2010.03.012

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). https://doi.org/10.1016/j.neuron.2004.10.014

Piazza, M., Mechelli, A., Price, C. J., & Butterworth, B. (2006). Exact and approximate judgements of visual and auditory numerosity: An fMRI study. *Brain Research*, *1106*(1), 177–188. https://doi.org/10.1016/j.brainres.2006.05.104

Piazza, M., Pica, P., Izard, V., Spelke, E. S., & Dehaene, S. (2013). Education Enhances the Acuity of the Nonverbal Approximate Number System. *Psychological Science*, *24*(6). https://doi.org/10.1177/0956797612464057

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). https://doi.org/10.1016/j.neuron.2006.11.022

Pica, P., Lemer, C., Izard, V., & Dehaene, S. (2004). Exact and approximate arithmetic in an Amazonian indigene group. *Science*, *306*(5695). https://doi.org/10.1126/science.1102085

Piéron, H. (1914). II. Recherches sur les lois de variation des temps de latence sensorielle en fonction des intensités excitatrices. *L'année Psychologique*, *20*, 17–96. https://doi.org/10.3406/psy.1913.4294

Pinel, P., & Dehaene, S. (2010). Beyond hemispheric dominance: Brain regions underlying the joint lateralization of language and arithmetic to the left hemisphere. *Journal of Cognitive Neuroscience*, *22*(1), 48–66. https://doi.org/10.1162/jocn.2009.21184

Pinheiro-Chagas, P., Wood, G., Knops, A., Krinzinger, H., Lonnemann, J., Starling-Alves, I., Willmes, K., & Haase, V. G. (2014). In how many ways is the approximate number system associated with exact calculation? *PLoS ONE*, *9*(11). https://doi.org/10.1371/journal.pone.0111155

Plaisier, M. A., Bergmann Tiest, W., & Kappers, A. M. L. (2009). One, two, three, many - Subitizing in active touch. *Acta Psychologica*. https://doi.org/10.1016/j.actpsy.2009.04.003

Pomè, A., Anobile, G., Cicchini, G. M., & Burr, D. C. (2019). Different reaction-times for subitizing, estimation, and texture. *Journal of Vision*. https://doi.org/10.1167/19.6.14

Pomè, A., Anobile, G., Cicchini, G. M., Scabia, A., & Burr, D. C. (2019). Higher attentional costs for numerosity estimation at high densities. *Attention, Perception, and Psychophysics*. https://doi.org/10.3758/s13414-019-01831-3

Pooresmaeili, A., Arrighi, R., Biagi, L., & Morrone, M. C. (2013). Blood oxygen level-dependent activation of the primary visual cortex predicts size adaptation illusion. *Journal of Neuroscience*, *33*(40). https://doi.org/10.1523/JNEUROSCI.1770-13.2013

Prado, J., Mutreja, R., & Booth, J. R. (2014). Developmental dissociation in the neural responses to simple multiplication and subtraction problems. *Developmental Science*, *17*(4), 537–552. https://doi.org/10.1111/desc.12140

Prado, J., Mutreja, R., Zhang, H., Mehta, R., Desroches, A. S., Minas, J. E., & Booth, J. R. (2011). Distinct representations of subtraction and multiplication in the neural systems for numerosity and language. *Human Brain Mapping*, *32*(11), 1932–1947. https://doi.org/10.1002/hbm.21159

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, *40*(3). https://doi.org/10.3758/BRM.40.3.879

Preacher, K. J., & Kelley, K. (2011). Effect size measures for mediation models: Quantitative strategies for communicating indirect effects. *Psychological Methods*, *16*(2). https://doi.org/10.1037/a0022658

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). https://doi.org/10.1016/j.actpsy.2012.02.008

Primi, C., Busdraghi, C., Tomasetto, C., Morsanyi, K., & Chiesi, F. (2014). Measuring math anxiety in Italian college and high school students: Validity, reliability and gender invariance of the Abbreviated Math Anxiety Scale (AMAS). *Learning and Individual Differences*, *34*. https://doi.org/10.1016/j.lindif.2014.05.012

Primi, C., Donati, M. A., Chiesi, F., & Morsanyi, K. (2018). Are there gender differences in cognitive reflection? Invariance and differences related to mathematics. *Thinking and Reasoning*, *24*(2).

https://doi.org/10.1080/13546783.2017.1387606

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

Railo, H., Koivisto, M., Revonsuo, A., & Hannula, M. M. (2008). The role of attention in subitizing. *Cognition*. https://doi.org/10.1016/j.cognition.2007.08.004

Räsänen, P., Salminen, J., Wilson, A. J., Aunio, P., & Dehaene, S. (2009). Computer-assisted intervention for children with low numeracy skills. *Cognitive Development*, *24*(4). https://doi.org/10.1016/j.cogdev.2009.09.003

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation*, *20*(4), 873–922. https://doi.org/10.1162/neco.2008.12-06-420

Reeve, R., Reynolds, F., Humberstone, J., & Butterworth, B. (2012). Stability and change in markers of core numerical competencies. *Journal of Experimental Psychology: General*, *141*(4). https://doi.org/10.1037/a0027520

Roggeman, C., Santens, S., Fias, W., & Verguts, T. (2011). Stages of nonsymbolic number processing in occipitoparietal cortex disentangled by fMRI adaptation. *Journal of Neuroscience*, *31*(19). https://doi.org/10.1523/JNEUROSCI.4503-10.2011

Roitman, J. D., Brannon, E. M., & Platt, M. L. (2007). Monotonic coding of numerosity in macaque lateral intraparietal area. *PLoS Biology*, *5*(8). https://doi.org/10.1371/journal.pbio.0050208

Rolison, J. J., Morsanyi, K., & O'Connor, P. A. (2016). Can i Count on Getting Better? Association between Math Anxiety and Poorer Understanding of Medical Risk Reductions. *Medical Decision Making*, *36*(7). https://doi.org/10.1177/0272989X15602000

Ross, J. (2003). Visual discrimination of number without counting. *Perception*. https://doi.org/10.1068/p5029

Ross, J., & Burr, D. (2012). Number, texture and crowding. In *Trends in Cognitive Sciences* (Vol. 16, Issue 4). https://doi.org/10.1016/j.tics.2012.01.010

Ross, J., & Burr, D. C. (2010). Vision senses number directly. *Journal of Vision*, *10*(2). https://doi.org/10.1167/10.2.10

Ross, J., & Speed, H. D. (1991). Contrast adaptation and contrast masking in human vision. *Proceedings of the Royal Society B: Biological Sciences*, *246*(1315). https://doi.org/10.1098/rspb.1991.0125

Rousselle, L., & Noël, M. P. (2007). Basic numerical skills in children with mathematics learning disabilities: A comparison of symbolic vs non-symbolic number magnitude processing. *Cognition*, *102*(3). https://doi.org/10.1016/j.cognition.2006.01.005

Sasanguie, D., De Smedt, B., Defever, E., & Reynvoet, B. (2012). Association between basic numerical abilities and mathematics achievement. *British Journal of Developmental Psychology*, *30*(2). https://doi.org/10.1111/j.2044-835X.2011.02048.x

Sasanguie, D., Defever, E., Maertens, B., & Reynvoet, B. (2014). The approximate number system is not predictive for symbolic number processing in kindergarteners. *Quarterly Journal of Experimental*

Psychology, *67*(2). https://doi.org/10.1080/17470218.2013.803581

Sasanguie, D., Göbel, S. M., Moll, K., Smets, K., & Reynvoet, B. (2013). Approximate number sense, symbolic number processing, or number-space mappings: What underlies mathematics achievement? *Journal of Experimental Child Psychology*, *114*(3). https://doi.org/10.1016/j.jecp.2012.10.012

Sawamura, H., Shima, K., & Tanji, J. (2002). Numerical representation for action in the parietal cortex of the monkey. *Nature*, *415*(6874), 918–922. https://doi.org/10.1038/415918a

Sawamura, H., Shima, K., & Tanji, J. (2010). Deficits in action selection based on numerical information after inactivation of the posterior parietal cortex in monkeys. *Journal of Neurophysiology*, *104*(2), 902–910. https://doi.org/10.1152/jn.01014.2009

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). https://doi.org/10.1111/desc.12372

Sereno, M. I., Pitzalis, S., & Martinez, A. (2001). Mapping of contralateral space in retinotopic coordinates by a parietal cortical area in humans. *Science*, *294*(5545). https://doi.org/10.1126/science.1063695

Shapiro, K. L., Arnell, K. M., & Raymond, J. E. (1997). The attentional blink. In *Trends in Cognitive Sciences*. https://doi.org/10.1016/S1364-6613(97)01094-2

Silver, M. A., Ress, D., & Heeger, D. J. (2005). Topographic maps of visual spatial attention in human parietal cortex. *Journal of Neurophysiology*, *94*(2). https://doi.org/10.1152/jn.01316.2004

Solomon, S. G., & Kohn, A. (2014). Moving sensory adaptation beyond suppressive effects in single neurons. In *Current Biology* (Vol. 24, Issue 20). https://doi.org/10.1016/j.cub.2014.09.001

Spielberger, C. D., Gonzalez, H. P., Taylor, C. J., Algaze, B., & Anton, W. D. (1978). Examination stress and test anxiety. *Stress Anxiety*, *5*, 167–191.

St. John-Saaltink, E., Kok, P., Lau, H. C., & De Lange, F. P. (2016). Serial dependence in perceptual decisions is reflected in activity patterns in primary visual cortex. *Journal of Neuroscience*. https://doi.org/10.1523/JNEUROSCI.4390-15.2016

Stanescu-Cosson, R., Pinel, P., Van De Moortele, P. F., Le Bihan, D., Cohen, L., & Dehaene, S. (2000). Understanding dissociations in dyscalculia: A brain imaging study of the impact of number size on the cerebral networks for exact and approximate calculation. *Brain*, *123*(11), 2240–2255. https://doi.org/10.1093/brain/123.11.2240

Starkey, G. S., & McCandliss, B. D. (2014). The emergence of “groupitizing” in children’s numerical cognition. *Journal of Experimental Child Psychology*, *126*, 120–137. https://doi.org/10.1016/j.jecp.2014.03.006

Starkey, P., Spelke, E. S., & Gelman, R. (1990). Numerical abstraction by human infants. *Cognition*, *36*(2). https://doi.org/10.1016/0010-0277(90)90001-Z

Starr, A., Libertus, M. E., & Brannon, E. M. (2013). Number sense in infancy predicts mathematical abilities in childhood. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(45).

https://doi.org/10.1073/pnas.1302751110

Steele, A., Karmiloff-Smith, A., Cornish, K., & Scerif, G. (2012). The Multiple Subfunctions of Attention: Differential Developmental Gateways to Literacy and Numeracy. *Child Development*, *83*(6). https://doi.org/10.1111/j.1467-8624.2012.01809.x

Suaŕez-Pellicioni, M., Núñez-Peña, M. I., & Colomé, À. (2013). Abnormal error monitoring in math-anxious individuals: Evidence from error-related brain potentials. *PLoS ONE*, *8*(11). https://doi.org/10.1371/journal.pone.0081143

Sullivan, J., Frank, M. C., & Barner, D. (2016). Intensive math training does not affect approximate number acuity: Evidence from a three-year longitudinal curriculum intervention. *Journal of Numerical Cognition*, *2*(2). https://doi.org/10.5964/jnc.v2i2.19

Swisher, J. D., Halko, M. A., Merabet, L. B., McMains, S. A., & Somers, D. C. (2007). Visual topography of human intraparietal sulcus. *Journal of Neuroscience*, *27*(20). https://doi.org/10.1523/JNEUROSCI.0991-07.2007

Taubert, J., Alais, D., & Burr, D. (2016). Different coding strategies for the perception of stable and changeable facial attributes. *Scientific Reports*. https://doi.org/10.1038/srep32239

Testolin, A., & McClelland, J. L. (2020). Do estimates of numerosity really adhere to Weber's law? A reexamination of two case studies. *Psychonomic Bulletin and Review*. https://doi.org/10.3758/s13423-020-01801-z

Thompson, P., & Burr, D. (2009). Visual aftereffects. *Current Biology*, *19*(1), R11-14. https://doi.org/10.1016/j.cub.2008.10.014

Tibber, M. S., Greenwood, J. A., & Dakin, S. C. (2012). Number and density discrimination rely on a common metric: Similar psychophysical effects of size, contrast, and divided attention. *Journal of Vision*, *12*(6). https://doi.org/10.1167/12.6.8

Togoli, I., Crollen, V., Arrighi, R., & Collignon, O. (2020). The shared numerical representation for action and perception develops independently from vision. *Cortex*, *129*. https://doi.org/10.1016/j.cortex.2020.05.004

Trommershäuser, J., Gepshtein, S., Maloney, L. T., Landy, M. S., & Banks, M. S. (2005). Optimal compensation for changes in task-relevant movement variability. *Journal of Neuroscience*. https://doi.org/10.1523/JNEUROSCI.1906-05.2005

Van Harskamp, N. J., & Cipolotti, L. (2001). Selective impairments for addition, subtraction and multiplication. Implications for the organisation of arithmetical facts. *Cortex*, *37*(3), 363–388. https://doi.org/10.1016/S0010-9452(08)70579-3

Vetter, P., Butterworth, B., & Bahrami, B. (2008). Modulating attentional load affects numerosity estimation: Evidence against a pre-attentive subitizing mechanism. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0003269

Vukovic, R. K., Kieffer, M. J., Bailey, S. P., & Harari, R. R. (2013). Mathematics anxiety in young children: Concurrent and longitudinal associations with mathematical performance. *Contemporary Educational Psychology*, *38*(1). https://doi.org/10.1016/j.cedpsych.2012.09.001

Wang, L., Mruczek, R. E. B., Arcaro, M. J., & Kastner, S. (2015). Probabilistic maps of visual topography in human cortex. *Cerebral Cortex*, *25*(10), 3911–3931. https://doi.org/10.1093/cercor/bhu277

Wang, Z., Hart, S. A., Kovas, Y., Lukowski, S., Soden, B., Thompson, L. A., Plomin, R., McLoughlin, G., Bartlett, C. W., Lyons, I. M., & Petrill, S. A. (2014). Who is afraid of math? Two sources of genetic variance for mathematical anxiety. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *55*(9). https://doi.org/10.1111/jcpp.12224

Wang, Z., Lukowski, S. L., Hart, S. A., Lyons, I. M., Thompson, L. A., Kovas, Y., Mazzocco, M. M. M., Plomin, R., & Petrill, S. A. (2015). Is Math Anxiety Always Bad for Math Learning? The Role of Math Motivation. *Psychological Science*, *26*(12). https://doi.org/10.1177/0956797615602471

Watanabe, S. (1998). Discrimination of “four” and “two” by pigeons. *Psychological Record*, *48*(3). https://doi.org/10.1007/BF03395279

Watson, A. B., & Pelli, D. G. (1983). Quest: A Bayesian adaptive psychometric method. *Perception & Psychophysics*, *33*(2). https://doi.org/10.3758/BF03202828

Wender, K. F., & Rothkegel, R. (2000). Subitizing and its subprocesses. *Psychological Research*, *64*(2), 81–92. https://doi.org/10.1007/s004260000021

Whalen, J., Gallistel, C. R., & Gelman, R. (1999a). Nonverbal Counting in Humans: The Psychophysics of Number Representation. *Psychological Science*. https://doi.org/10.1111/1467-9280.00120

Whalen, J., Gallistel, C. R., & Gelman, R. (1999b). Nonverbal Counting in Humans: The Psychophysics of Number Representation. *Psychological Science*, *10*(2). https://doi.org/10.1111/1467-9280.00120

Whalen, J., McCloskey, M., Lesser, R. P., & Gordon, B. (1997). Localizing arithmetic processes in the brain: Evidence from a transient deficit during cortical stimulation. *Journal of Cognitive Neuroscience*, *9*(3). https://doi.org/10.1162/jocn.1997.9.3.409

Wilson, A., Dehaene, S., Pinel, P., Revkin, S., Cohen, L., & Cohen, D. (2006). Principles underlying the design of “The Number Race”, an adaptive computer game for remediation of dyscalculia. *Behavioral and Brain Functions*, *2*(1). https://doi.org/10.1186/1744-9081-2-19

Wilson, A. J., Dehaene, S., Dubois, O., & Fayol, M. (2009). Effects of an adaptive game intervention on accessing number sense in low-socioeconomic-status kindergarten children. *Mind, Brain, and Education*, *3*(4). https://doi.org/10.1111/j.1751-228X.2009.01075.x

Wilson, A. J., Revkin, S. K., Cohen, D., Cohen, L., & Dehaene, S. (2006). An open trial assessment of “the number race”, an adaptive computer game for remediation of dyscalculia. *Behavioral and Brain Functions*, *2*. https://doi.org/10.1186/1744-9081-2-20

Wynn, K. (1992). Addition and subtraction by human infants. *Nature*, *358*(6389). https://doi.org/10.1038/358749a0

Xia, Y., Leib, A. Y., & Whitney, D. (2016). Serial dependence in the perception of attractiveness. *Journal of Vision*. https://doi.org/10.1167/16.15.28

Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition*, *74*(1). https://doi.org/10.1016/S0010-0277(99)00066-9

Xu, X., & Liu, C. (2008). Can subitizing survive the attentional blink? An ERP study. *Neuroscience Letters*. https://doi.org/10.1016/j.neulet.2008.05.063

Young, C. B., Wu, S. S., & Menon, V. (2012). The Neurodevelopmental Basis of

Math Anxiety. *Psychological Science*, *23*(5). https://doi.org/10.1177/0956797611429134

Zago, L., Petit, L., Turbelin, M. R., Andersson, F., Vigneau, M., & Tzourio-Mazoyer, N. (2008). How verbal and spatial manipulation networks contribute to calculation: An fMRI study. *Neuropsychologia*, *46*(9). https://doi.org/10.1016/j.neuropsychologia.2008.03.001

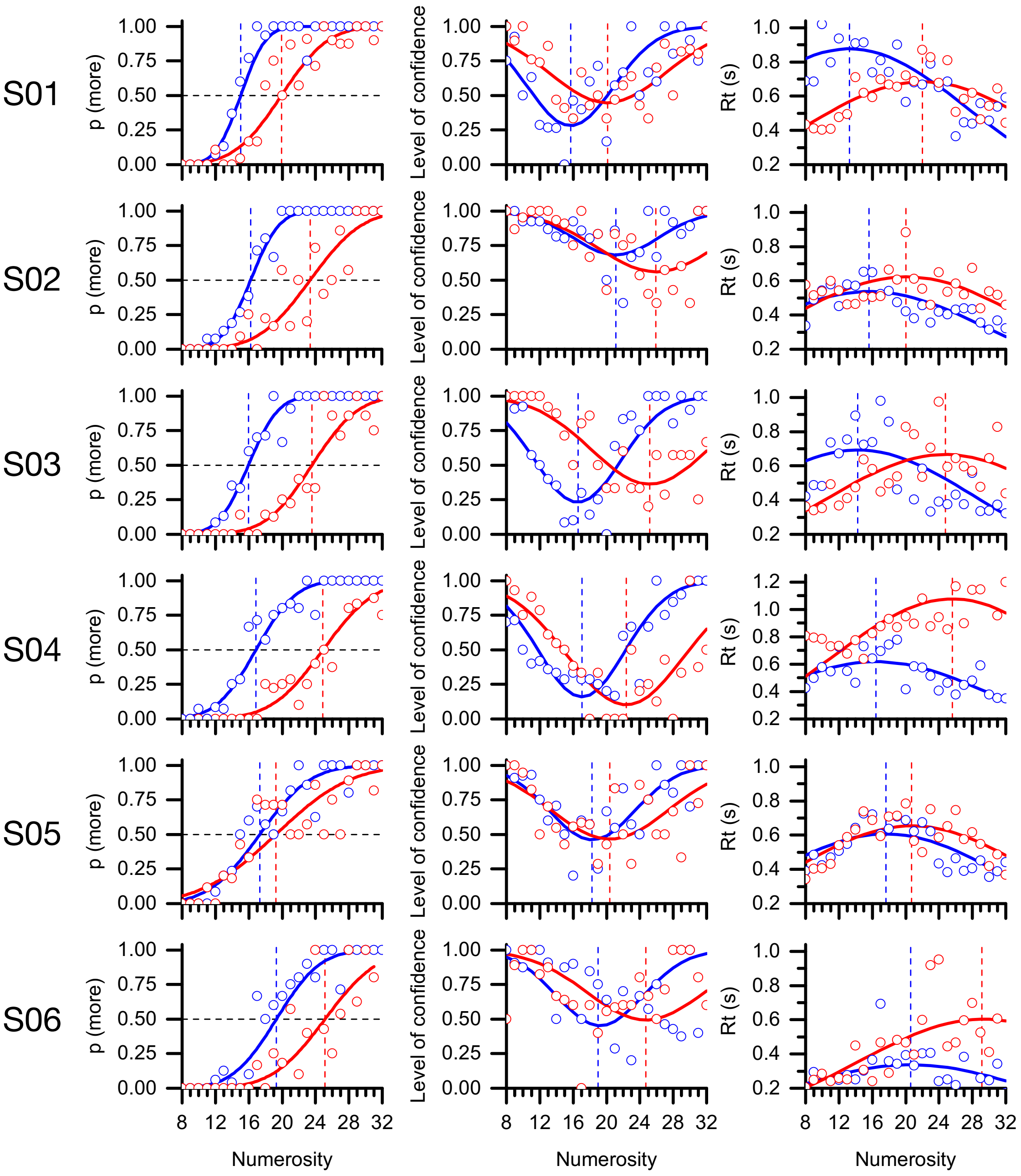
Zhou, X., Chen, C., Zang, Y., Dong, Q., Chen, C., Qiao, S., & Gong, Q. (2007). Dissociated brain organization for single-digit addition and multiplication. *NeuroImage*, *35*(2), 871–880. https://doi.org/10.1016/j.neuroimage.2006.12.017

Appendix

Appendix

Results of individual analysis from study in chapter 2

Visual condition



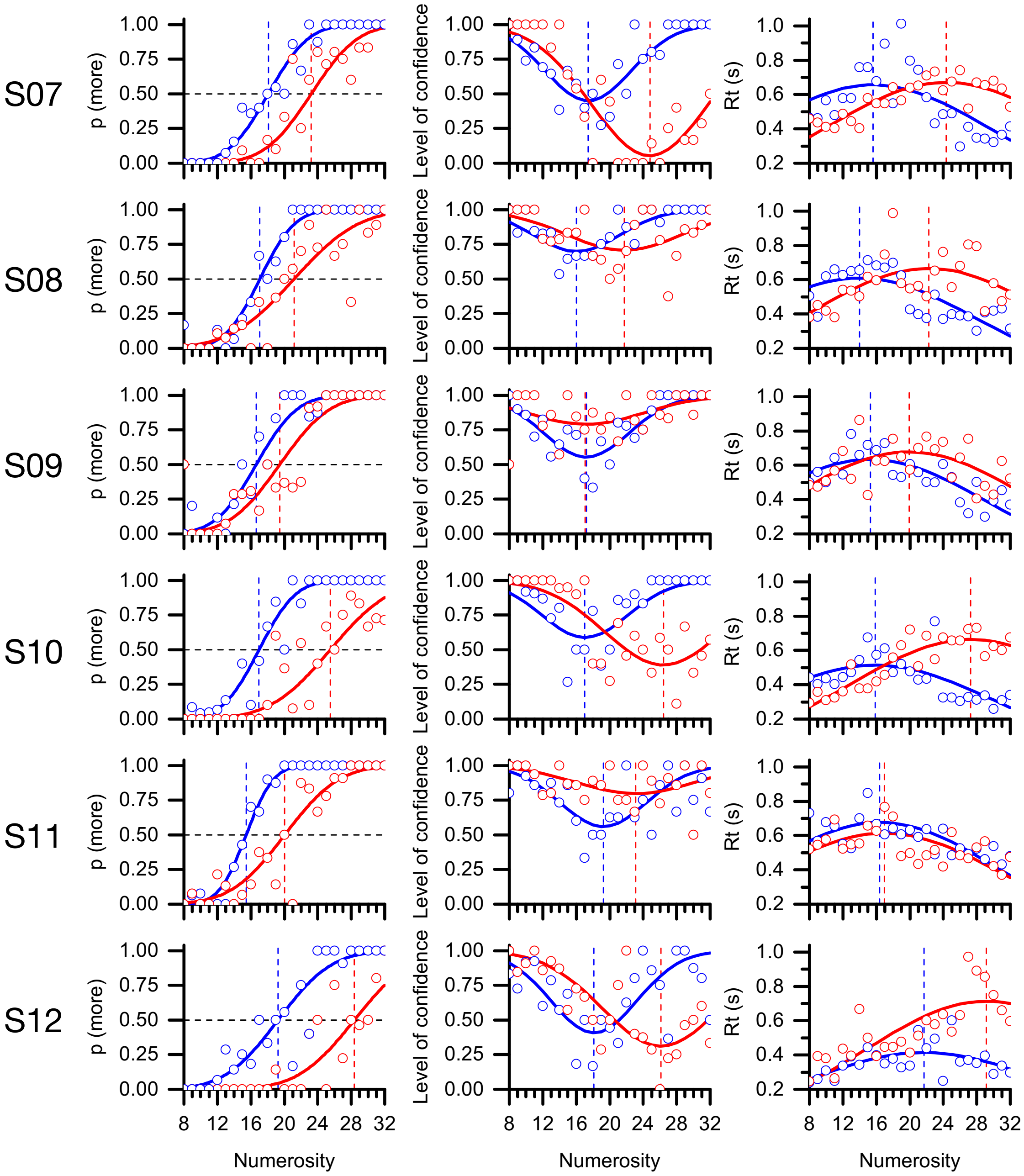
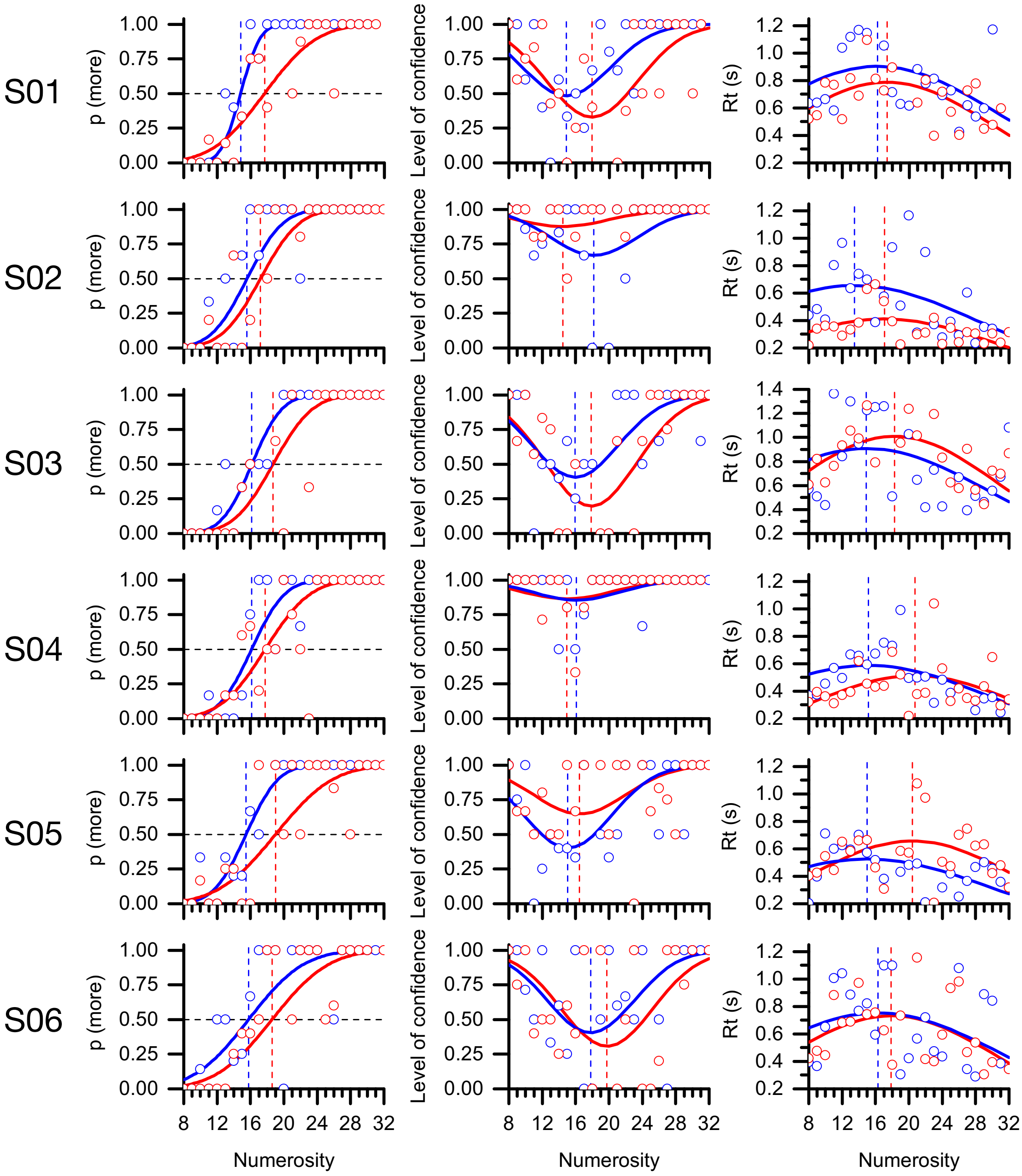


Figure A1. Single-participant plots for the visual condition. The first column plots psychophysical functions of proportion of trials when the test was seen as more numerous than the neutral probe, as a function of physical numerosity (baseline in blue and adapt to high in red). The second column shows average confidence, the third mean reaction-times as a function of physical numerosity. Each row represents a participant. The dashed lines show the PSEs or the peaks of the best-fit gaussians to the confidence or reaction time distributions.

Motor condition



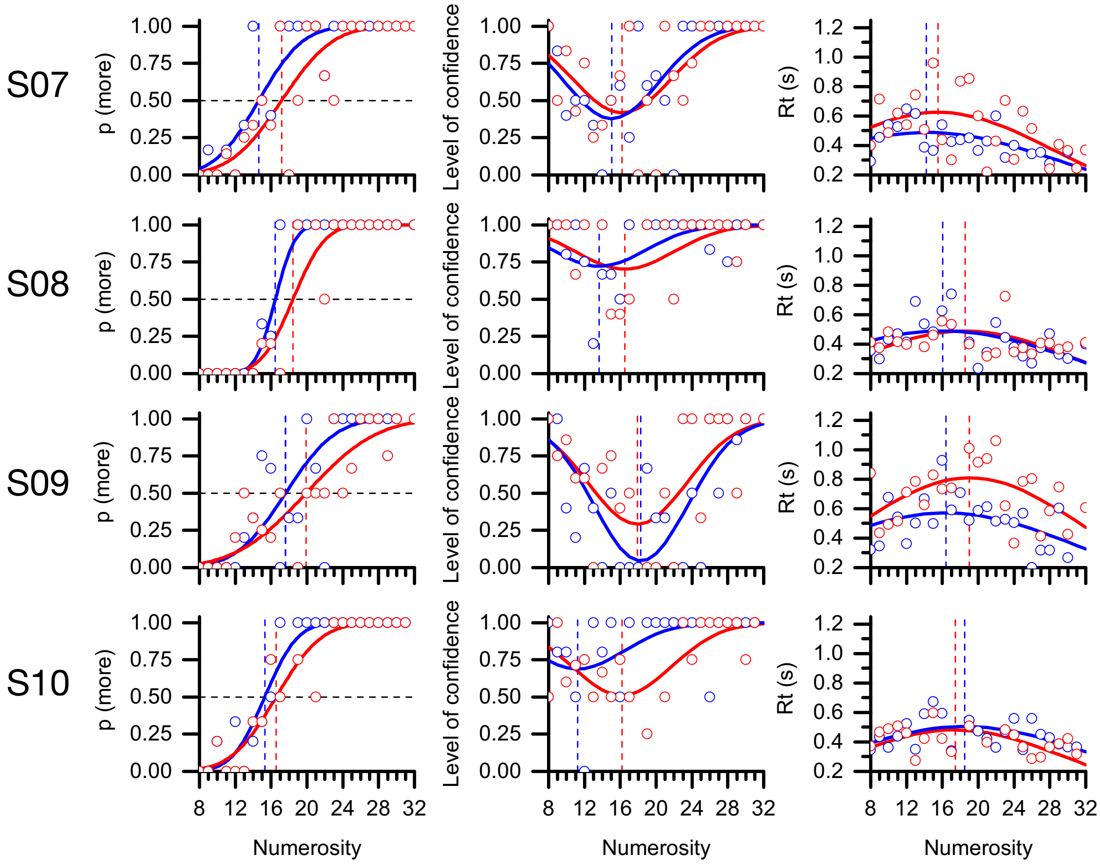
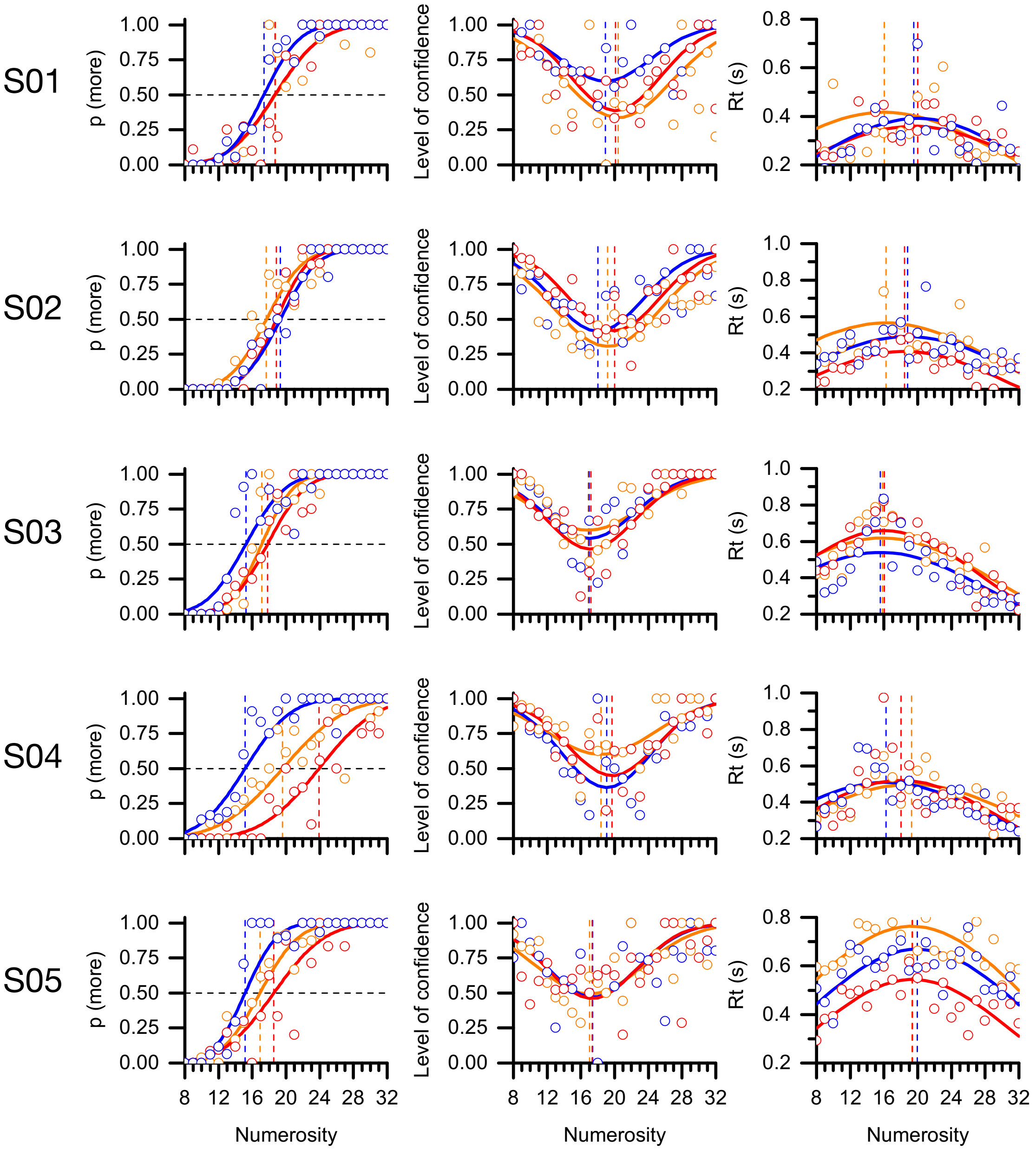


Figure A2. Single-participant plots for the motor condition. Conventions as for Figure A1, except that here blue refers to slow tapping adaptation and red to fast tapping adaptation.

Control condition



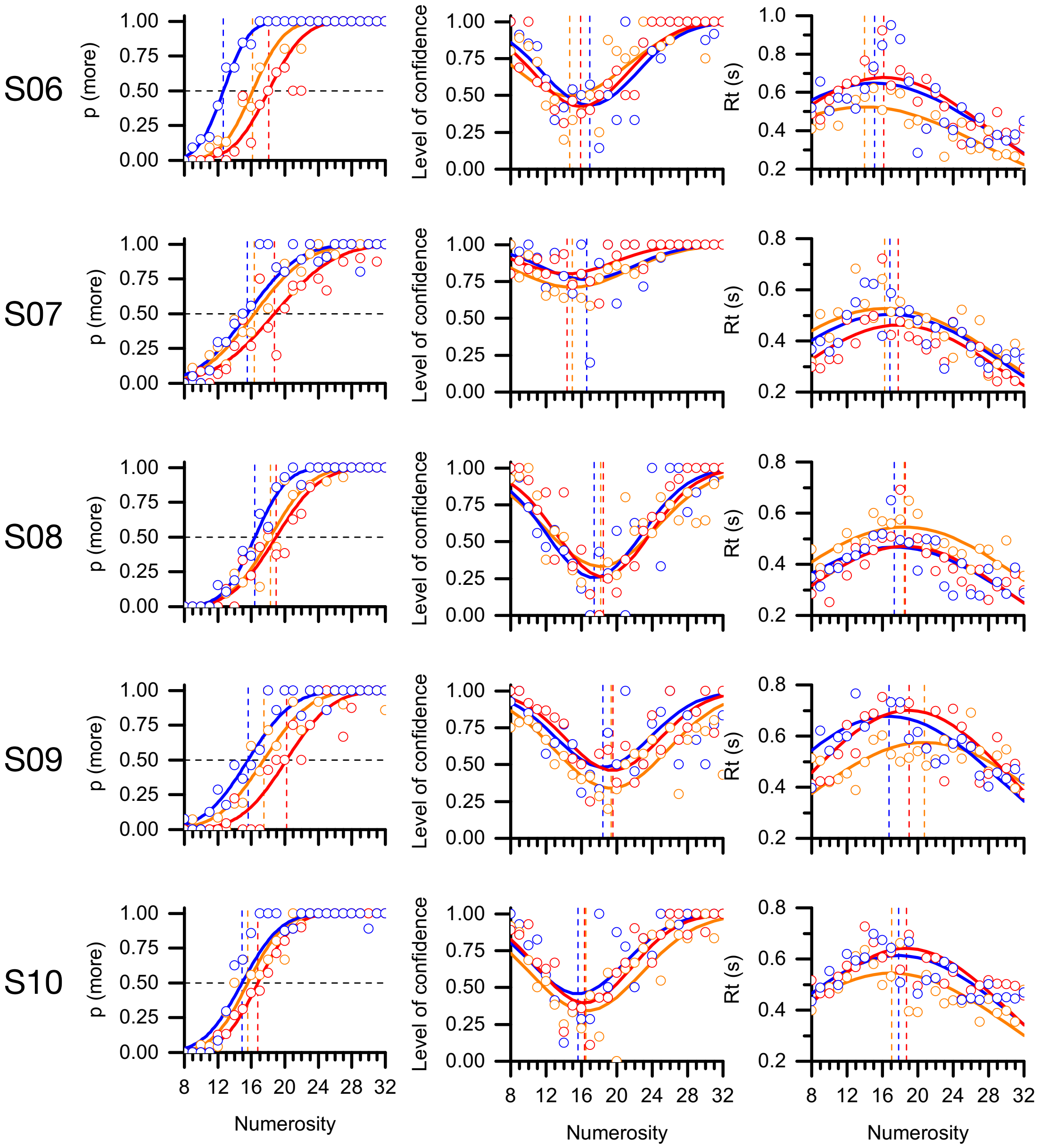


Figure A3. Single-participant plots for the control condition. Conventions as for Figure A1, except that here orange refers to baseline, blue to the “reward-low” condition and red to the “reward-high” condition.

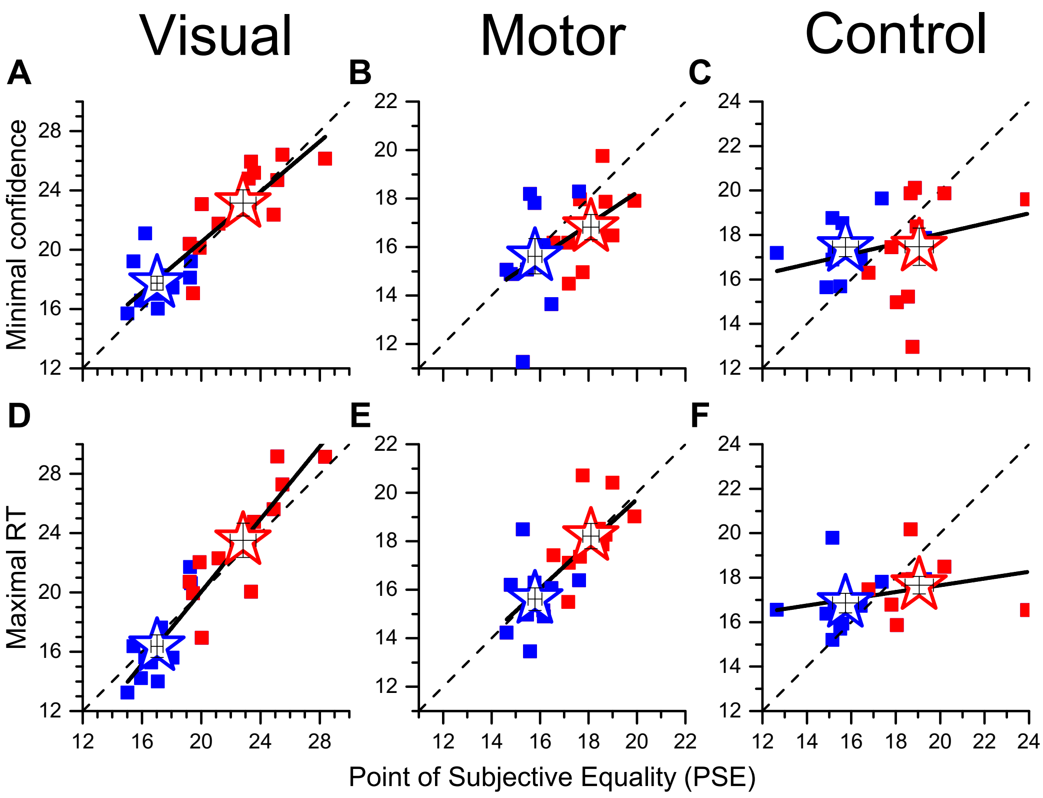


Figure A4. Minima of confidence distributions (A, B, C) and maxima of reaction-time distributions (D, E, F) plotted against PSEs for the visual (A-D), motor (B-E) and control condition (C-F). Single subject data are shown as filled squares and group averages as large open stars. Black lines represent best linear regression lines. Slopes are close to unity in all the adaptation conditions (0.85 and 1.22 for confidence and reaction time for visual adaptation (A-D); 0.66 and 0.92 for confidence and reaction time for motor adaptation (B-E)). Slopes in the control condition are close to zero (0.23 and 0.15 for confidence and reaction time (C-F)).

Cluster summary table of univariate analysis presented in chapter 5

Ungrouped numerosity> Ungrouped shape control								
Left Hemisphere								
Cluster	MNI X	MNI Y	MNI Z	Max	CWP	Size	NVtxs	Label
4	-38.5	-43.3	37.5	3.679	0.04586	315.79	849	S_postcentral
Right Hemisphere								
Cluster	MNI X	MNI Y	MNI Z	Max	CWP	Size	NVtxs	Label
3	27.2	-60.4	37.1	3.423	0.00020	1094.08	2110	S_intrapariet_and_P_trans
4	44.2	-46.5	36.1	4.597	0.00020	954.12	2469	S_interm_prim-Jensen
5	8.5	24.1	39.2	3.558	0.00459	387.63	741	G_and_S_cingul-Mid-Ant
6	50.0	3.5	21.6	3.766	0.00499	383.61	799	S_precentral-inf-part
Ungrouped shape control> Ungrouped numerosity								
Left Hemisphere								
Cluster	MNI X	MNI Y	MNI Z	Max	CWP	Size	NVtxs	Label
1	-17.7	36.5	44.0	-4.859	0.00020	2832.55	4941	G_front_sup
2	-40.0	-66.4	29.4	-5.507	0.00020	1308.62	2737	S_temporal_sup
3	-35.0	22.9	-16.6	-4.767	0.00958	408.80	828	G_orbital
Right Hemisphere								
Cluster	MNI X	MNI Y	MNI Z	Max	CWP	Size	NVtxs	Label
1	12.9	58.9	20.0	-4.883	0.00020	2467.94	3865	G_front_sup
2	45.7	-64.4	29.4	-5.340	0.00020	1404.88	2758	G_pariet_inf-Angular

Grouped numerosity>Grouped shape control								
Left Hemisphere								
Cluster	MNI X	MNI Y	MNI Z	Max	CWP	Size	NVtxs	Label
5	-45.0	-40.6	40.4	4.019	0.0002	779.20	2018	S_postcentral
6	-45.7	1.6	28.4	3.785	0.0002	646.48	1405	S_precentral-inf-part
7	-40.6	38.3	4.0	4.351	0.0002	585.19	808	S_front_inf
8	-31.1	-69.1	36.8	5.242	0.00040	486.05	950	S_intrapariet_and_P_trans
Right Hemisphere								
Cluster	MNI X	MNI Y	MNI Z	Max	CWP	Size	NVtxs	Label
1	38.1	-49.1	37.7	6.479	0.00020	2622.14	5961	S_intrapariet_and_P_trans
6	41.8	3.4	26.1	4.497	0.00020	731.80	1426	S_precentral-inf-part
8	8.3	30.1	39.9	5.130	0.00020	511.91	974	G_front_sup
9	26.3	5.3	47.2	3.256	0.00659	372.08	767	S_front_sup
12	30.3	20.1	6.9	4.196	0.01732	325.15	835	S_circular_insula_sup
13	40.8	32.4	10.4	3.460	0.01792	324.11	547	S_front_inf
14	55.7	-52.2	-12.6	3.226	0.02030	319.22	461	G_temporal_inf
Grouped shape control> Grouped numerosity								
Left Hemisphere								
Cluster	MNI X	MNI Y	MNI Z	Max	CWP	Size	NVtxs	Label
1	-16.3	39.2	45.2	-5.341	0.00020	2808.72	4870	G_front_sup
2	-59.0	-51.7	24.8	-6.97	0.00020	1769.03	3758	G_pariet_inf-Supramar
3	-9.2	-48.0	7.2	-5.408	0.00020	1563.27	2682	G_cingul-Post-ventral
4	-13.5	-48.2	40.4	-4.749	0.00020	925.68	2007	S_subparietal
9	-57.5	2.4	8.7	-4.261	0.01415	343.65	885	G_and_S_subcentral

10	-47.5	-33.1	20.5	-5.407	0.01494	338.37	708	G_pariet_inf-Supramar
11	-23.6	-42.7	54.9	-4.178	0.03351	293.79	608	S_postcentral
Right Hemisphere								
Cluster	MNI X	MNI Y	MNI Z	Max	CWP	Size	NVtxs	Label
2	15.5	44.2	2.8	-5.151	0.00020	1463.08	2403	G_and_S_cingul-Ant
3	48.7	-60.3	24.7	-4.157	0.00020	1440.89	2931	G_pariet_inf-Angular
4	14.5	-83.0	35.3	-4.405	0.00020	998.63	1416	G_cuneus
5	22.5	50.5	29.8	-4.180	0.00020	827.71	1394	G_front_middle
7	7.4	-56.9	50.9	-5.218	0.00020	569.32	1604	G_precuneus
10	12.3	-67.3	-1.7	-3.496	0.00838	354.70	449	G_oc-temp_med-Lingual
11	11.2	-10.3	40.5	-3.656	0.01137	340.48	879	G_and_S_cingul-Mid-Post
15	30.1	-41.9	-8.5	-3.785	0.03901	290.10	543	S_oc-temp_med_and_Lingual

(Grouped numerosity>Grouped shape control)∩(Ungrouped numerosity> Ungrouped shape control)							
Left Hemisphere							
Cluster	MNI X	MNI Y	MNI Z	Max	Size	NVtxs	Label
1	-45.0	-40.6	40.4	4.019	276.69	742	S_postcentral
Right Hemisphere							
Cluster	MNI X	MNI Y	MNI Z	Max	Size	NVtxs	Label
1	38.1	-49.1	37.7	6.479	900.72	2330	S_intrapariet_and_P_trans
2	32.3	-65.9	28.0	5.111	899.90	1737	S_oc_sup_and_transversal
3	8.3	30.1	39.9	5.130	371.29	706	G_front_sup
4	41.8	3.4	26.1	4.497	365.03	768	S_precentral-inf-part

<table>
<tr><td colspan="9">(Grouped numerosity>Grouped shape control)>(Ungrouped numerosity> Ungrouped shape control)</td></tr>
<tr><td colspan="9">Left Hemisphere</td></tr>
<tr><td>Cluster</td><td>MNI X</td><td>MNI Y</td><td>MNI Z</td><td>Max</td><td>CWP</td><td>Size</td><td>NVtxs</td><td>Label</td></tr>
<tr><td>1</td><td>-37.5</td><td>48.5</td><td>-1.0</td><td>4.845</td><td>0.01057</td><td>360.30</td><td>509</td><td>S_orbital_lateral</td></tr>
<tr><td>2</td><td>-38.1</td><td>-68.0</td><td>46.1</td><td>3.746</td><td>0.04215</td><td>281.61</td><td>495</td><td>G_pariet_inf-Angular</td></tr>
<tr><td colspan="9">Right Hemisphere</td></tr>
<tr><td>Cluster</td><td>MNI X</td><td>MNI Y</td><td>MNI Z</td><td>Max</td><td>CWP</td><td>Size</td><td>NVtxs</td><td>Label</td></tr>
<tr><td>1</td><td>36.0</td><td>-70.7</td><td>43.4</td><td>4.349</td><td>0.00499</td><td>382.25</td><td>753</td><td>G_pariet_inf-Angular</td></tr>
</table>

A5 Table 1. Cluster summary tables. For each contrast and for the conjunction analysis displayed in Figure 5.2 and 5.3, the table reports: the cluster number (Cluster), the MNI coordinates of the maximally activated vertex within each cluster (MNI X, Y, Z), the maximum -log10(p-value) in the cluster (Max), the cluster-wise p-value of each cluster (CWP), the cluster surface area in mm2 (Size), the number of vertices for each cluster (NVtxs) and the name of the region (as defined by the Destrieux Atlas) containing the maximally activated vertex within a given cluster (label).