

MÁSTERES de la UAM

Facultad de Medicina / 16-17

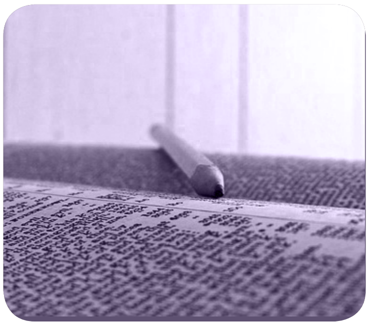
Neurociencia

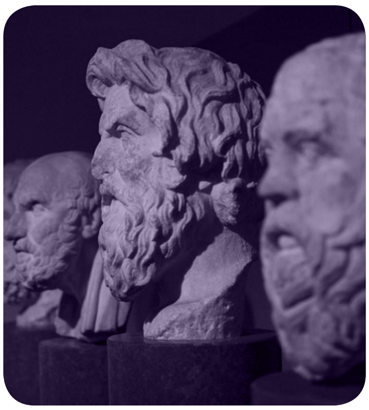
UNIVERSIDAD AUTONOMA DE MADRID

Campus Internacional excelencia UAM CSIC+

Exploring the role of visual cues and calibration in the numerical estimation task

Inés Abalo Rodríguez





UAM EDICIONES

TRABAJO FIN DE MÁSTER

EXPLORING THE ROLE OF VISUAL CUES AND CALIBRATION IN THE NUMERICAL ESTIMATION TASK

Autora: **Inés Abalo Rodríguez**

Director: Simone Cutini
Co-Directora: AlmudenaCapilla
Afiliación: Cognition and Language Laboratory (CoLab), Università degli Studi di Padova.

Madrid, 15 de junio de 2017

Simone Cutini
Department of Developmental Psychology, University of Padova
Via Venezia 8, 35131, Padova, Italy

Almudena Capilla
Departamento de Psicología Biológica y de la Salud (Facultad de Psicología, UAM).
C/ Iván Pavlov, 6, 28049, Madrid, España.

Como Directores del Trabajo Fin de Máster titulado *"EXPLORING THE ROLE OF VISUAL CUES AND CALIBRATION IN THE NUMERICAL ESTIMATION TASK"* que ha realizado bajo nuestra supervisión Inés Abalo Rodríguez como parte de sus trabajos para lo obtención del Título de Máster en Neurociencia por la Universidad Autónoma de Madrid,

DECLARAMOS:

Que reúne los méritos suficientes para poder ser presentado y defendido públicamente.

Madrid, 13 de junio de 2017

Fdo.: SimoneCutini

Fdo.: Almudena Capilla

INDEX

SUMMARY

In our daily life, we need to face numbers in multiple situations. During the last decades, it has been believed that there were two systems in order to process numerosity: one for the small numerosities (Object Tracking System) and one for large numerosities (Approximate Number System).

In this sense, ANS is thought to be a system that allows us to estimate large numerical quantities and that constitutes a core system from which the numerical thinking and the mathematical abilities are developed. According to neuroimaging studies, it depends on the activity of the intraparietal sulcus, whose neurons would be sensitive to specific numerosities.

However, during the last years, the constant increase of controversial results in this topic highlighted severe methodological issues, and the impact of visual cues constitute one of the most relevant problems that need to be tackled. The criticism revolved around the influence of visual cues pushed some scientists to reconsider the theoretical framework of numerosity processing: indeed, a Sensory Integration System, entirely based on visual cues, has been recently proposed as an alternative to the Approximate Number System.

In order to better understand these aspects, the current work investigates for the first time the influence that visual cues have when performing the calibration in an estimation task. In this task, participants are asked to estimate the numerosity of arrays of dots with time pressure. Before doing so, they *calibrate* their performance: before they start estimating numerosities, they are presented a set of dots whose numerosity is said. In that way, calibration creates a better mapping between the stimuli and the numerical label. Thanks to this, ANS is thought to be measured in a more precise way.

Our inter-subjects study carried out an estimation task with three different calibration images. All of them had the same numerosity but differed in terms of their visual cues. According to the traditional theories, no difference should be found between the groups given that calibration is a process entirely based on numerosity. However, our results showed that visual cues have a relevant influence in the calibration.

In conclusion, the current study evidences that a reconsideration of the traditional theories about calibration seems to be mandatory, given the impact of visual cues. In this respect, the Sensory Integration theory might be an interesting framework to consider.

Key words: ANS, numerosity, mathematics, visual cues, estimation task, calibration, Sensory-Integration System.

INTRODUCTION

1. INTRODUCTION

In our daily life, we have to face numbers in several situations: when we work on computers, when we play some sports or even we go shopping, numbers happen to be an essential part of these activities. But how are humans able to deal with them?

During the last twenty five years it has been suggested that in order to understand the concept of symbolic number, human beings rely on two different innate non-symbolic number systems: the Object Tracking System (OTS) and the Approximate Number System (ANS) (Cantlon & Brannon, 2006). Both of them are suggested to be present across different human societies as well as other animal species (Cantlon, 2012; Cantlon et al., 2009). These two system are considered to be two "core number systems" that allow human beings to develop the sense of numerosity, beside being foundational for the skills requested in more complex mathematics. (Feigenson, Dehaene & Spelke, 2004; Piazza, 2010; Halberda & Feigenson, 2008; Halberda, Mazzocco & Feigenson, 2008).

In the present work we will be focusing on one of these two systems: the ANS. In order to understand the current work, the main concepts related to the topic will be explained as well as the most relevant theories that are needed to comprehend the current issues in this research field.

First of all, we will shortly introduce the two core number systems that were already mentioned and their main differences will be explained. Then, we will focus on ANS, and we will explain the characteristics of this ancient system in detail. We will describe its features, its neural substrate, its evolutionary importance, its foundational role in mathematics, the mechanisms that support it and the tasks traditionally used to measure it.

Afterwards, the methodological problems when assessing ANS will be exposed, with particular attention to the issue of visual cues. We will then introduce a new theory that has been proposed as an alternative to ANS explanation to describe number processing (Gebuis et al., 2016). As it will be explained, this new theory posits an essential role to visual cues throughout the process. Finally, we will briefly revise the existing algorithms to create the stimuli in order to justify the choices adopted for the current experiment.

2. TWO SYSTEMS TO PROCESS NUMEROSITY

As previously announced, it is believed that the capacity of the human being to deal and process numerosity relies on two innate systems: the Object Tracking System (OTS) and the Approximate Number System (ANS). The features of each system as well as their mechanisms have been proved to be different, so researchers agree in describing two different systems (Feigenson, Dehaene &Spelke, 2004; Piazza et al., 2011; Revkin et al., 2008; Cutini et al., 2014.

2.1. OBJECT TRACKING SYSTEM (OTS)

The Object Tracking System (OTS) is one of the two "core number system" that the human being is equipped with and it processes the exact representation of a small number of objects (smaller than 4 or 5 items) (Feigenson, Dehaene & Spelke, 2004).

It was already in 1871, when the logician W. S. Jevons observed that, when trying to quickly estimate the numerosity of a handful of beans thrown on a plate, his performance was errorless for up to four units. The ability of quickly, accurately and effortlessly enumerate sets up to four objects was named "subitizing" (Kaufman, Lord, Reese, & Volkmann, 1949) from the Latin word “*subitus*”, meaning immediate.

The Object Tracking System seems to be the cognitive source of subitizing because it is thought to be a system that allows to represent small amount of objects as separate entities and track them in space and time (Piazza, 2010). This system is based on some concepts like cohesion, continuity and contact, which allow both animals and humans to perceive the boundaries of the objects and predict if an object will move, stay still or interact with another object (Spelke & Kindler, 2007).

As we mentioned before, the limit of this system is set in approximate four elements tracked at the same time. It is believed that this limitation is due to the relationship that this system has with visual short-term memory and attention (Alvarez & Cavanagh, 2004; Scholl, 2001; Burr, Turi & Anobile, 2010). This storage capacity of the OTS increases during the development (Ross-Sheey, Oakes & Luck, 2003) and it is an element of relevant individual differences (Halberda, Mazzocco & Feigenson 2008).

Finally, it is necessary to mention that the neural substrates of the OTS seems to be associated with regions in the posterior parietal cortex and in the occipital cortex, similar to the brain regions underlying visual memory and attention processes (Todd & Marois, 2004; Xu & Chun, 2006).

2.2. APPROXIMATE NUMBER SYSTEM (ANS)

On the other hand, Approximate Number System (ANS) processes large numerosities (beyond 4 or 5 items): we are able to estimate large numerosities in a very short time with little expense of effort and energy. Unlike performance in the subitizing range, the estimation is far from being errorless: the error percentage increases as the number of elements of the stimulus increases (Jevons, 1871).

This system, which will be described in detail in the next section, seems to follow the Weber's law (Ross, 2003): the threshold of discrimination (intended as the smallest difference noticeable) between two stimuli increases linearly as the intensity of the stimulus increases (Fechner, 1860).

3. ANS

3.1. WHAT IS ANS?

As mentioned in the previous section, ANS is one of the two innate non-symbolic number system that human brain is equipped with in order to deal with the environment demands (Cantlon & Brannon, 2006). More specifically, ANS is responsible for the processing of large numbers of objects (larger than 4 or 5 items).

The ANS system is usually conceptualized as a logarithmically compressed number line that represents numerosities as partially overlapping Gaussian curves (Dehaene, 2003). In that way, the neurons that conform this system fire to specific numerosities. The probability of a specific neuron fire is represented in a Gaussian curve, because the system is intrinsically approximate. For instance, a neuron that fires for numerosity 10 will fire maximally when 10 items are presented, but it will be also firing at numerosities 8 and 12 (although at a slower firing rate). This is the reason why it is said that the Gaussian curves are overlapping.

It is important to notice that this overlap increases logarithmically: the larger the numerosities, the larger the overlap. This leads to two effects well known and studied in the field:

- The **distance effect** (Moyer & Landauer, 1967). Given that neural representations overlap, it will be easier to compare two numbers that are further apart than two numbers closer to each other. (See ***Figure 1***, A versus B).

- The **size effect** (Moyer &Landauer, 1967). As already mentioned, the overlap increases with increasing number. Therefore, it will be easier to compare two relatively small numbers than two relatively large numbers, even if there is the same distance between them. (See ***Figure 1***, A versus C).

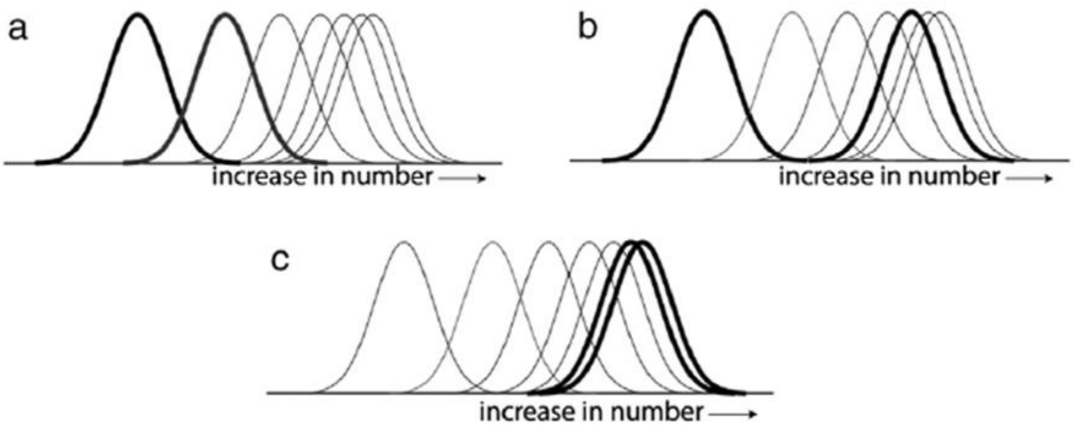


Figure 1. *A schematic representation of the neural representation of number (adapted from Gebuis et al., 2016). Distance effect (comparing A versus B) and size effect (comparing A versus C) can be observed.*

Because of this last effect, it can be said that the ability to discriminate between two numerosities is best described by the relative difference between both numerosities, using for instance the ratio (Piazza et al., 2010). There are several behavioral and imaging studies that suggest that this ratio dependence is the core mechanism of the ANS system (Cantlon et al., 2006; Piazza et al., 2004; Roggeman et al., 2011).

3.2.NEURAL SUBSTRATE

The neural substrate of the ANS has been shown in neuroimaging studies, and it is more clearly defined than the one of OTS. In that way, these studies consistently indicate regions in the intraparietal sulcus as neural correlate of the ANS both in children and adults (Piazza & Izard, 2009; Cutini et al. 2014).

Additionally, the previous mentioned effects have been also shown in single cell recordings in monkeys. For instance, Nieder & Merten (2007) documented neurons that were maximally responsive to their preferred number. Apart from that, they showed that even if these neurons also responded to the neighboring numbers, there was a decrease of the activation when increasing the distance from the preferred number.

3.3. AN ANCIENT SYSTEM

The ANS theory proposes that the ANS is an evolutionarily ancient system shared across all the cultures (e.g., Pica et al., 2004) and different species (e. g., Agrillo, 2015; Beran et al., 2015; Cantlon et al., 2009; Dehaene, 1997; Hauser et al., 2003; Piazza et al., 2004). It appears logical that, as well as it is useful for our specie, animals might also need an ability to discriminate or estimate numerosity. For instance, it would be beneficial for fishes to join the largest shoal as this decreases the chance of being caught by a predator (Agrillo et al., 2008).

Because of these reasons, animals such as bears, monkeys, chicks, fish and pigeons have been pointed out as species that present a spontaneous representation of numerosity (e.g. Dadda et al., 2009; Agrillo et al., 2007; Agrillo et al., 2010; Agrillo et al., 2012; Emmerton & Renner, 2009; Hauser et al., 2000; Rugani et al., 2014; Vonk & Beran, 2012; for a review: Brannon & Reitman, 2003). Even without intensive training sessions, animals could made comparisons of items and respond to simple arithmetic problems (e.g. Agrillo et al., 2010; Cantlon & Brannon, 2005;Flombaum et al., 2005; Hauser et al., 2003).

Despite the consistency of these results, this kind of studies have been strongly criticized recently, as it will be explained later: it could be possible that animals follow different strategies (based, for example, on visual cues) instead of numerosity to perform in these tasks given that there has not been a strict control of these variables in these studies (Gómez-Laplaza & Gerlai, 2013).

3.4. ANS AND MATHEMATICS

The main reason that explains the importance of Aproximate Number System in humans is the foundational role that the ANS (more than the OTS) has in the acquisition of complex numerical concepts (Piazza, 2010).

According to current theories of cognitive development, knowledge acquisition is based on a limited set of "core knowledge" systems over which cultural learning occurs. Those systems reside in cerebral circuits that have been selected throughout evolution because of their relevant functions. These circuits are plastic enough for acquiring new functions. These systems are also called "neurocognitive start-up tools" and, according to the literature, ANS would be the one that allows the numerical development in general (Butterworth, 2005; Dehaene, 2009; Piazza, 2010), as well as more advanced mathematics in particular (Halberda & Feigenson, 2008).

There are several findings in support of these claims. First, there are already signs of ANS existence in infants of 6 months (Xu and Spelke, 2000). Second, ANS precision increases with increasing age until young adulthood (Brannon, Suanda, & Libertus, 2007; Halberda & Feigenson, 2008; Wood & Spelke, 2005; Xu, Spelke, & Goddard, 2005); however, this increment is impaired in children with arithmetic deficiencies (Piazza et al., 2010).

Furthermore, there are several studies that have shown a relationship between the ANS and more complex arithmetic skills (e.g. Halberda et al., 2008; Inglis et al., 2011; Mazzocco et al., 2011; Starr et al., 2013a,b), which have made think to some researchers that developing programs to train this capacity at schools could help to increase performance in mathematics (Park & Brannon, 2013; Piazza et al., 2013; Libertus et al., 2013)

Despite these facts, as it will be explained later, some critics have been made to this relationship given that other studies have not found the mentioned correlation (see section 4).

3.5. MECHANISMS OF ANS

In this section it will be briefly explained the mechanism that it is believed to be underneath the ANS, even if it is not fully understood.

The most influential models suggest that the process takes place through four different stages (see ***Figure 2***). Firstly, there is a *sensory stage* in which the incoming sensory properties are processed, such as the size of the elements, the convex hull, the density, etc. (see the section 5 for a description of these features). Visual properties are confounded with numerosity given that they increase or decrease when increasing or decreasing the numerosity. This is why there is a second step called *normalization stage* in which these sensory cues are removed and pure numerosity is retained.

At this point, numerosity information feeds to the accumulator in the third step (*accumulator stage*). This step is usually pictured as a non accurate breaker that pours water into a cylinder for each item perceived (Dehaene, 1997). In this sense, each time the water is poured the total amount of water becomes increasingly inaccurate, which results in an approximate and not accurate estimate of numerosity (*read-out stage*). In this sense, the bigger the number of scoops of water is, the more inaccurate the final response is. This is consistent with the observation that numerosity discrimination performance decreases with increasing numerosities (Gebuis & Reynvoet, 2012c; Izard & Dehaene, 2008).

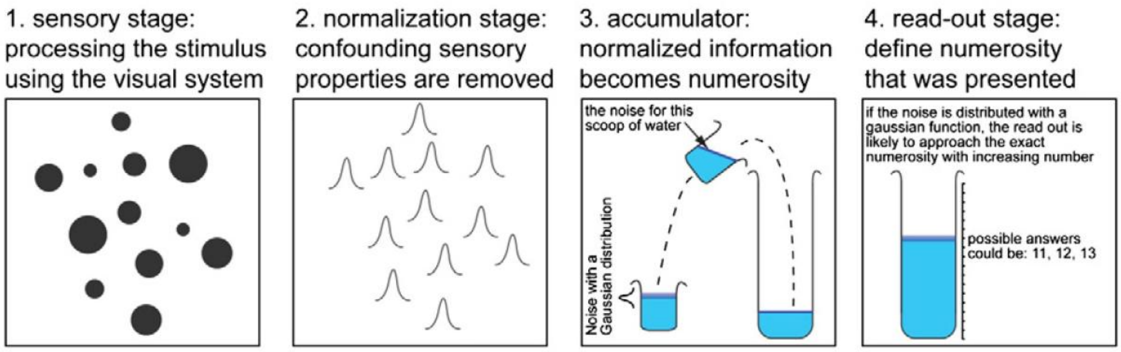


Figure 2.*The accumulator model that explains the mechanism underneath the Approximate Number System (from Gebuis et al., 2016).*

3.6. <u>HOW IS ANS MEASURED</u>

In order to measure the Approximate Number System, different tasks have been traditionally used, being the most popular the *comparison task* and, in second place, the *estimation task.* Both of them require presenting a set of stimuli (normally arrays of dots) for a very brief time in order to avoid subjects to have time to count.

a) ***<u>Comparison task</u>***

The comparison tasks is the most popular task when measuring ANS. Two images of dots are shown to the subjects just for some milliseconds. After this brief time, they are asked to choose the one that, according to their judge, had a bigger quantity of dots.

The way the images are presented varies from one experiment to another (which might be leading to the some contradictory results that are found). In some cases, subjects are simply shown two different images from which they have to choose one. In others, they are presented one single image with two different colors dots, so they have to choose the color whose dots were more numerous.

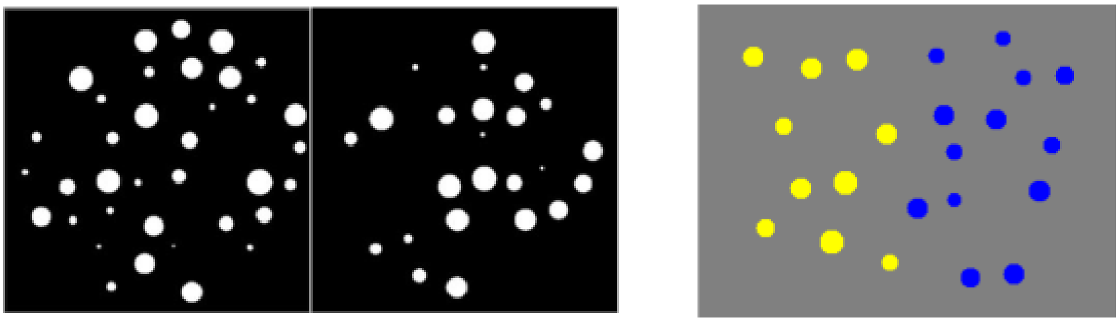


Figure 3. *On the left, a pair of stimulus generated for a comparison task (Gebuis & Reynvoet, 2012b). On the right, a different type of stimulus generated also for a comparison task (Halberda & Ly, 2013).*

b) ***Estimation task***

In the estimation task, subjects are asked to estimate the numerosity of one single image that contains an array of dots. In that way, once the image is briefly shown, they are asked to say how many dots they estimate they have just seen. Subjects response can be either vocal or written.

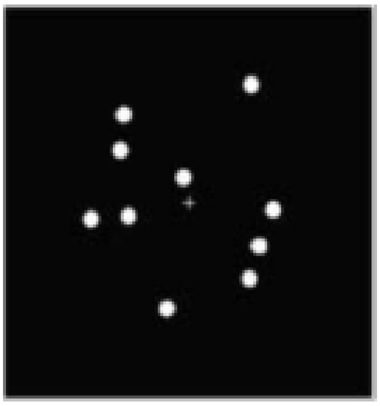


Figure 4.*A stimulus generated for an estimation task (Dehaene et al., 2005).*

Compared to comparison tasks, estimation tasks elicit a very poor performance. In order to improve subjects' accuracy, Izard & Dehaene (2008) suggested to implement a process called "**calibration**". In little words, calibration consists in presenting to the subjects a stimulus while saying to them the numerosity that the array of dots has. In that way, subjects are able to map the analog representations of numerosity and the verbal numerical labels.

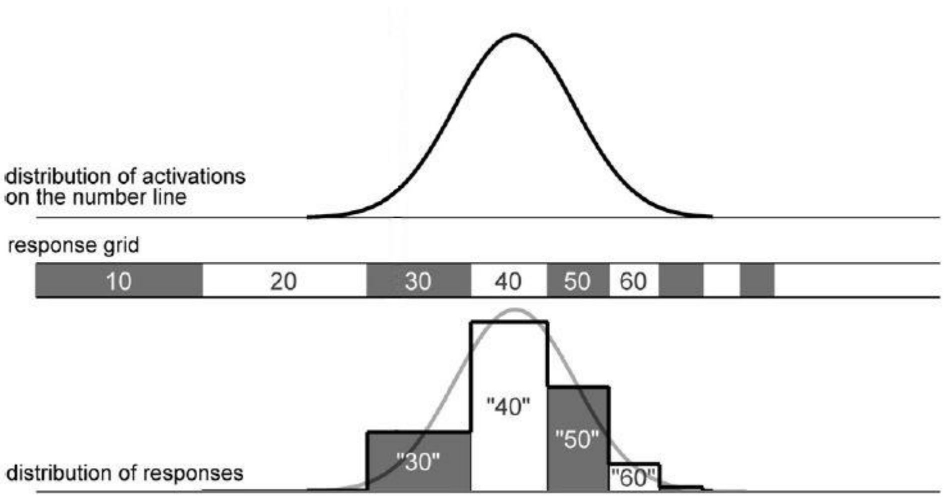


Figure 5.*Model of numerosity process suggested by Izard & Dehaene (adapted from Izard & Dehaene, 2008).*

On their view, calibration helps because it stretches or compresses the response grid that subjects had. According to their model, each perceived numerosity is encoded on an internal continuum (the *number line*) that it is logarithmically scaled. Hence, when a numerosity is perceived, it generates an activation on the number line that it assumed to be Gaussian. The correspondence between the number line and the verbal numerical response is done through the response grid: the number lines is divided in several segments, each of them associated with a different verbal label (Dehaene & Mehler, 1992). The calibration would be able to stretch or compress this grid, so a better mapping between the numerosity and the verbal label would be possible.

Apart from this, it is necessary to mention that the authors saw that the influence of the calibration was not restricted to the neighborhood of the calibrated numerosity; instead, it extended to the whole range of numerosities tested. In this sense, a simple calibration at the beginning of the experiment would be enough to improve subjects answers.

4. METHODOLOGICAL PROBLEMS WHEN STUDYING ANS

The research on the Approximate Number System has not been exempt of problems and criticisms given the amount of contradictory results that are found in the field.

Despite the facts described in the previous section that seemed to evidence a relationship between non-symbolic number representation and arithmetic (see section 3.4), some recent works hypothesized that this relationship might be spurious (e.g., Gebuis and Reynvoet, 2015).

Some studies do find a relationship between non-symbolic numerosity performance and arithmetic (e.g. Halberda et al., 2008; Libertus et al., 2013; Inglis et al., 2011; Mazzocco et al., 2011; see section 3.4 of the current work) whereas others do not (De Smedt & Gilmore, 2011; Gilmore et al., 2013; Rousselle & Noel, 2007; Sasanguie, De Smedt, Defever, & Reynvoet, 2012; Haist, Wazny, Toomarian & Adamo, 2015; Newton, Waring, & Penner-Wilger, 2014).

How can these contradictory results be explained? Most of the researchers have attributed this divergence to methodological differences across studies (see for further methodological explanations: Feigenson et al., 2013; Guillaume et al., 2015; Price et al., 2012; Smets et al., 2014). In fact, there is a big deal of variation between studies in terms of numerosities, ratios, time of exposition, type of task, indexing or stimuli used. Dietrich et al. (2015) pointed this variability out as a relevant problem on their review, and provided some guidelines to follow in order to avoid this fact and achieve a bigger consistence across ANS experiments.

Apart from these methodological variability, some authors have indicated other problems in ANS methodology: the indexes and tasks used to describe and measure ANS that are treated as equivalent might not be in fact equivalent (Lindskog et al., 2013; Price et al., 2012).

First of all, a recent study (Inglis et al., 2014) tried to understand the causes of the opposing results found. In order to do that, they studied in detail the four different indexes that are used to describe ANS and that are assumed to measure similar properties. They concluded that three out of the four indexes (between which it is included the Weber fraction, the most used index when describing ANS) had really poor and different statistical properties and, hence, should not be used anymore nor treated as equivalent.

Regarding to the variability in tasks, a couple of studies have measured ANS with different tasks that are used in research and that are treated like equivalent across the experiments. Specifically, they compared *non symbolic large numerosity comparison* (comparison task with large numerosities), *non symbolic small-numerosity comparison* (comparison task with small numerosities), *symbolic large numerosity comparison* (comparison task made with digits representing large numerosities) *symbolic small numerosity comparison* (comparison task made with digits representing small numerosities), *non symbolic addition* (participants had to say if the third and last of the dot arrays presented was approximately the sum of the two previous ones) and *symbolic* *addition* (identical to the previous one but numerosities were presented as Arabic numerals). Surprisingly, no correlation emerged between the tasks, neither in the ones for adults (Gilmore et al., 2011) nor in the ones for children (Gilmore et al., 2015).

Furthermore, another group (Guillaume et al., 2015) have recently found no correlation between the two most used tasks (estimation task and comparison task). In fact, comparison task have been strongly criticized lately as a measurement of ANS: it is considered to be also measuring visual cues when they are not controlled (Szűcs et al., 2013; visual cues problem will be detailed later) and to be measuring also inhibition when they are controlled (Clayton & Gilmore, 2014). In other words, the most used paradigm to measure ANS happens to be highly inaccurate.

Reached this point, the methodological problem when studying ANS is more than evident and needs to be seriously considered for future research in the field. But is it the only issue that explains the big variability in the results found? According to Gebuis et al. (2016) it is not: there is also a theoretical problem that need to be considered. In this recent paper, they have proposed a new theory that explains the numerosity process. According to their view, it would not be the ANS, but in fact a "sensory-integration system" (SIT) the one in charge of estimating numerosities.

At this point we can finally introduce the "visual cues problem", which has been a relevant topic in ANS field during these last years. Nowadays, most of the researchers acknowledge the importance of visual cues as a relevant variable to control when designing these paradigms. However, visual cues are the feature in which SIT is based, which concedes them not only a methodological role but also a theoretical one. Because of this reason, we have decided to analyze them in a different section.

5. THE PROBLEM OF THE VISUAL CUES

As already announced, in this section we will be revising the "visual cues problem". Nowadays "visual cues problem" is understood as a methodological issue by most of the researchers, but some of them have started to consider it a theoretical problem instead.

5.1. WHAT ARE "VISUAL CUES"?

Sensory properties can be defined as each non-numerical cue that comprises numerosity stimuli (Gebuis et al., 2016). When talking about the visual domain, we can rename them as "visual cues", understanding for them features as the density of the image, the diameter of the dots or the total surface that they occupy.

These features happen to be relevant when studying numerosity processes because of the commonly high correlation between numerosity and visual cues: when numerosity increases, the visual cues of the image (like its density or the total surface that the stimuli occupy) also increase.

This high correlation can be observed easily in daily life. For instance, if we compare five apples to twenty apples in real life, the "amount of red" (total surface) will be also larger in the more numerous set of apples. In this sense, this high correlation can lead according to some authors to confound both notions (Gebuis et al., 2016). In the previous example, there is no certainty of the person's answer being based on the numerosity per se, given it could be based instead on the high-correlated visual cue.

For this reason, every research group that aims to investigate the non-symbolic numerical cognition needs to face this natural correlation when generating the stimuli.

5.2. WHICH ARE THE MOST RELEVANT VISUAL CUES?

There are five visual cues that have been studied in more detailed in the literature: *total surface, total contour length, average diameter, convex hull* and *density*.

- **Total surface:** the aggregate surface of all the dots included in a single stimulus. The value is obtained by summing the areas of each dot. In some works, this concept is referred to as “total luminance” or “aggregate surface”.
- **Total contour length:** the sum of the perimeters of all the dots.
- **Average diameter:** the average diameter of the dots included in a single stimulus.

It is important to notice that total surface, total contour length and average diameter are very difficult to disentangle, given that manipulating one has an effect on the others.

- **Convex hull: “**the smallest contour that includes all of the dots of a single stimulus as if an elastic band were wrapped around the dots” (Gebuis & Reynvoet, 2011). In some papers, this concept is referred to as “area extended”.
- **Density:** the concept of density is not consistent across the non-symbolic numerical cognition literature. Although the concept of density is often described as “inter-item spacing” (Gebuis & Reynvoet, 2011; Santens et al., 2010), the two formulas

equivalently applied by the large majority of research groups only take into account features related to the array of dots and not to the single dots (density = convex hull / total surface or density = total surface / convex hull). Hence, they provide a very broad approximation of the idea of inter-item spacing. However, the paradigm developed by De Marco et al., (in prep.), which was used in the current experiment, calculates the density value as the average distance between dots.

5.3. IMPORTANCE OF VISUAL CUES THROUGHOUT THE YEARS

The influence of visual properties of the stimulus when studying non-numerical judgments is, besides intuitive, noticed for decades. In the 70's, Ginsburg (1978) found out that the way an array of dots is arranged (regular or random), that is a density manipulation, influenced participant's estimation (under or overestimation).

Even though, visual cues did not start to be seriously taken into account until the beginning of the new century. From this date on, researchers have conceded more importance to the control of visual properties when studying numerosity.

In this attempt of taking visual cues into account, we can broadly distinguish two streams of research: the one that attempts to *control* the sensory cues and the one that tries to *manipulate* sensory cues (Gebuis et al., 2016).

a) ***Controlling visual cues***

After realizing the importance that visual cues had in non-symbolical numerosity processing, some studies have tried to ***control*** the sensory cues to investigate numerosity processing independently of them(e.g. Xu & Spelke, 2000; Burr & Ross, 2008; Cantlon et al., 2006; Izard et al., 2009; Piazza et al., 2004; Roggeman et al., 2011). They attempt to balance the visual properties of the stimuli across the different numerosities presented to participants. In other words, the sensory cues are manipulated so that each single sensory cue was not helpful to judge numerosity throughout the experiment (Gebuis et al., 2016).

The idea behind this approach is that the consistent correlation between visual properties and numerosity might trigger a secondary strategy in which participants disregard numerosity in favor of the visual properties. According to this approach, if the visual properties are balanced across numerosities (e.g., the size of the dots is not consistently correlated with numerosity, no matter if negatively or positively), it is possible to measure the pure numerosity judgment, assuming that the results are not influenced in any way by the visual properties of the stimuli. In this sense, visual cues are viewed as a methodological problem that can be corrected.

However, studies aiming to control sensory cues neglect the fact that a perfect control for sensory cues is impossible: it is not possible to create two set of items with exactly the same sensory cues while differing in numerosity (Gebuis et al., 2016). Hence, the visual cues control is necessarily limited to the control of some of the properties at a time, so it cannot be concluded that participants do not rely on *any* sensory cue. Because of these reasons, the results of those experiments could be attributed to both numerosity processes or to the confounding effect of sensory cues.

b) *Manipulating visual cues*

Opposing to this approach, some studies try to, instead of controlling, *manipulate* sensory cues to investigate their role in numerosity processes. (e.g. Dakin et al., 2012; Gebuis & Reynvoet, 2012a; Szűcs & Soltesz, 2008; Szűcs et al., 2007).

In these studies, visual cues of sets of dots that represent the same numerosity are manipulated. Therefore, if participant's decisions vary when the visual cues change but the numerosity remains constant, it can be concluded that numerosity judgments depends to some extent to sensory properties.

In that way, some of the latest papers studied the influence of visual cues in estimating numerosities, finding out that this influence is noticeable both in comparison tasks (Szűcs et al., 2013) and estimation tasks (Gebius & Reynvoet, 2012c). Hence, instead of trying to remove their influence, visual cues are integrated as an inevitable part in number processing. In this respect, these researchers do not consider visual cues as a purely methodological problem that can be controlled, but rather as an inevitable element of the process that cannot be removed (which implies a change in the theoretical approach in numerosity process).

All these last evidences have recently led to a new theoretical framework that, instead of ignoring visual cues or giving them a secondary role, as the traditional view of ANS does, proposes a sensory-integration process in which the numerosity estimation process is carried out through the sensory properties of the stimuli (Gebuis et al., 2016).

6. SIT: A NEW APPROACH TO UNDERSTAND NUMEROSITY

Gebuis et al. (2016) have recently proposed a new framework to understand numerosity process based on the increasing importance that visual cues were receiving. Given the strong relationship between numerosity and visual cues in real life, they proposed that it is more likely to process the numerosity through the visual cues than operating independently from them.

To this end, they first review the current literature of ANS in a critical way questioning its existence (that is, questioning the existence of an innate system enabling to perceive pure numerosity). First of all, they consider as invalid the results of all the papers that did not have a proper control of the visual cues given that they do not have to be necessarily explained by numerosity-but could be instead explained by the visual cues-. They also criticize the poor methodology of the multi-modal experiments and suggest that further research is needed with these paradigms. Furthermore, they strongly criticize the lack of ecological validity of the experiments trying to *control* visual cues, explaining that the situations are artificial and not representative of the real contexts, as well as the problems and artifacts derived by these attempts of control (e. g., comparison tasks measuring also inhibition, Clayton & Gilmore, 2014).

Regarding the evolutionary view of ANS, that claims it existence in other species, the authors explain that given the natural relationship between visual cues and numerosity, the behavior of animals could be explained through the first ones (e. g. Gómez-Laplaza & Gerlai, 2013). In other words, animals could be answering, instead to numerosity (e.g. "which tree has a bigger number of red fruits"), to the visual cues highly correlated to it (e.g. "which tree has a bigger surface of red").

This point reached, they propose a Sensory-Integration Theory (from now on, SIT) as an alternative explanation for the numerosity process. This theory has two main differences compared to the ANS traditional theory: *the role of sensory properties* and *the output of the numerosity processing*.

First of all, for the Sensory Integration Theory visual cues are not something secondary that have a mere influence: according to this theory, the numerosity process relies on the sensory properties present on the stimuli. In few words, they propose that different sensory cues are simultaneously taken into account to come to an integrated representation of number (Gebuis et al., 2016).

Because of this reason, there is nothing similar to the normalization stage of the ANS: given that visual cues are considered to be the element through which the process is carried out, there is not an attempt of removing them from the process.

In second place, the output for the Sensory Integration Theory is not an abstract number, at least not according to the traditional definition: "*Adults can be said to rely on an abstract representation of number if their behavior depends only on the size of the number involved, not on the specific verbal or non-verbal means of denoting them*" (Dehaene et al., 1998 p.356; See also McCloskey, 1992, p. 497 for a similar definition). Given that numerosity processing happens through the visual cues and there is no normalization stage, there is not a moment in which people relied on an "abstract representation" purely based on the size of the number.

In that way, and contrary to the ANS theory, the output of the Sensory Integration Theory is more likely to be a sensory representation of number that allows estimating which set of items is numerically larger.

6.1. COMPARISON AND ESTIMATION ACCORDING TO SIT

Focusing now on comparison and estimation tasks, this model predicts that performance in both tasks would be different given that their requirements are also different, which explains the lack of correlation between them (Guillaume et al., 2015).

In that way, subjects would perform well in comparison tasks according to SIT, given that sensory cues of the two arrays of dots can be weighted and compared. However, the performance in the estimation tasks will be much worse because in these tasks sensory cues are not helpful. This fact explains why the indexes calculated from estimation tasks are always lower than the ones calculated from comparison tasks.

In other words, given that the system described by the Sensory Integration Theory relies on visual cues, it is likely that performance would be better on the tasks in which visual cues are helpful. As previously explained, in comparison tasks subjects are asked to select the image that has more dots. Given that numerosity correlates with visual cues, and in case the numerosity processing would be carried out through the visual cues, as SIT suggests, a good performance would be expected as subjects would be balancing the different visual properties in order to give an answer.

On the other hand, estimation tasks require a numerical answer to be given. In order to do that, visual cues are not helpful, so in case the numerosity process was based on the visual cues, the performance on this kind of tasks would be expected to be worse compared to the comparison tasks. This is exactly what is found in literature.

6.2. CALIBRATION IN ESTIMATION ACCORDING TO SIT

As previously explained, performance is quite poor in estimation tasks. In order to improve subjects' accuracy, calibration has been suggested: through calibrating the mapping between the analog representations of numerosity and the verbal numerical labels, participants' accuracy in estimation tasks is improved (Izard & Dehaene, 2008).

According to SIT, this improvement is based on the visual cues. The authors of the theory explain that showing a reference image provides subjects the opportunity to establish a relationship between the size of the different sensory cues and numerosity. In that way, they say, once this relation has been established, each following stimulus can be compared on the basis of the sensory cues of the former. Hence, if the visual cues of the new stimuli suggest that it has more items, then the numerosity estimate is increased accordingly.

7. HOW TO CREATE THE STIMULI

In order to create the stimuli that are used for non-symbolic numerical experiment, researchers have been using different programs. In all the literature, there had been only three algorithms that take visual cues into account. We will now revise them briefly in order to justify the choice we made for the current experiment.

7.1. TRADITIONAL ALGORITHMS

Since visual cues appeared to be relevant when estimating and comparing non-symbolical numerosities, two algorithms that took them into account were developed.

a) ***Dehaene, Izard and Piazza (2005)***

Dehaene, Izard and Piazza (2005) developed the first algorithm capable of generating different sets that respected specific constraints. Their Matlab script generates a set divided into two subsets: half of the stimuli keep total surface constant across numerosities and the other half keep average size of the dots constant across numerosities. In this way, it is assumed that participants are discourage to base their judgments on the visual cues of the stimuli because if they do it, their performance will be negatively affected considering that no single visual cue is predictive for numerosity throughout all trials (Smets, Sasanguie, Szücs & Reynvoet, 2015).

The script developed by Dehaene et al. (2005) -or related variants of this method-, has been used by many research groups (e.g. Halberda & Feigenson, 2008; Halberda et al., 2012).

However, this method has been strongly criticized given that it takes for granted that controlling one single visual cue at a time is enough to eliminate its influence. In this sense, it would be an insufficient method if people integrate information from multiple visual cues (e.g., if they relied on both diameter and aggregate surface). This method would be classified inside of the "controlling visual cues" category (Gebuis et al., 2016; see section 5.3).

b) <u>Gebuis and Reynvoet (2011)</u>

As an attempt to deal in a more appropriate way with visual cues, Gebuis & Reynvoet (2011) developed a new algorithm for generating sets of non-symbolic number stimuli. This program can be used for both the comparison task (Gebuis & Reynvoet, 2012b) and the estimation task (Gebuis & Reynvoet, 2012a), as well as for priming and habituation (Gebuis & Reynvoet, 2011).

Their Matlab function controlled two different visual cues: *convex hull* and *diameter*. Both of them were larger in half of the stimuli with larger numerosity (positive correlation with numerosity) while, in the other half, larger visual cues were associated with the smaller numerosity (negative correlation with numerosity). In that way, it was believed that visual cues were controlled, due to the fact that, in the whole set, they did not consistently correlate with numerosity (half of them held a negative correlation).

Consequently, this algorithm supersede the limitations for which Dehaene et al. method was strongly criticized. In fact, the relationship between these two algorithms have been investigated. Smets et al. (2015) confronted the results of the non-symbolic number comparison task using a fully congruent set (no control), a set generated with Dehaene et al. (2005) method (single feature manipulation) and a set generated with Gebuis & Reynvoet (2011) method (multiple features manipulation). The results of this experiment prove that the ANS accuracy and Weber’s fractions are consistently higher if measured using the fully congruent and the Dehaene’s set, meaning that a set generated with more restrictive constraints on visual features (Gebuis & Reynvoet, 2011), hence adding less visual cues to the numerosity information, actually leads the participants to a worse performance. Because of this reason, they concluded that Dehaene's method was obsolete and that all the papers carried out which such a method did not have a proper control of the visual cues.

Despite these facts, this way of controlling visual cues is not problems free, at least not when applied to comparison tasks. In fact, the negative correlation between numerosity and visual cues have transformed this test into a *Stroop like* task. Participants have to discriminate along a task relevant numerical dimension while they are supposed to neglect task-irrelevant non-numerical parameters (visual cues) of displays (Szűcs et al., 2013). Because of this, inhibition becomes a skill required in the task which leads to problems, as previously explained (see section 4; Clayton& Gilmore, 2014).

7.2. A NEW AND MORE SUITABLE ALGORITHM

a) ***De Marco et al. (in prep.)***

We designed this new algorithm as an improvement of some of the features that Gebuis & Reynvoet (2011) had and so it was chosen for the current experiment: for the conditions that our study had, this program was the one that best fitted our purpose and let us develop the idea with more flexibility.

In this sense, De Marco, Cutini et Abalo (2017) method offers a list of advantages that makes it the most suitable algorithm, at least for the experiment here described. In particular, the advantages more relevant for our choice are:

- **Customization**. This algorithm do not have predetermined built-in conditions, which means that it provides a big flexibility when designing the experimental paradigm.
- **Exact control.** With respect to the previous methods that controlled the visual properties of the overall sets trough equating the average values, the new algorithm offers the possibility of controlling the visual parameters in an exact way, on a stimulus by stimulus basis.
- **Density measure**. This algorithm provides a new way of measuring density: the average distance between dots. According to their authors, even if this measure is more difficult to obtain, it is more pertinent to the definition that is often given to density ("Inter-item spacing"). In this sense, density has in the paradigm the opposite sign compared to the traditional measure: the denser an image, the smaller the value of "density" is according to this new algorithm (because the denser an image, the shortest the average distance between dots is).

- **Contour length**. This visual cue had been progressively abandoned over the years in research given its correlation with total surface. However, even if both parameters are highly correlated, this correlation is not linear. Hence, different manipulations in these two different visual cues can lead to pretty different appearances of the stimuli. Because of this reason and due to its relevance, this new algorithm includes a control of this visual cue.

8. HYPOTHESIS OF OUR EXPERIMENT

In this experiment we aimed to assess the importance that the visual cues have in the calibration of the estimation task, in a way it had never studied before. As previously explained, this method was suggested in order to improve subjects performance when performing an estimation task (Izard & Dehaene, 2008). However, the mechanism through which the improvement is achieved remains unclear. Izard & Dehaene (see section 3.6.b) explained it through a better mapping between the response grid and the number line. On the other hand, Gebuis at al. (2016) hypothesized that, according to their theory, the improvement would be due to the relationship established between the visual cues and the numerical label (see section 6.2).

In order to help to better understand the mechanism underneath the calibration, we decided to examine the influence that visual cues have in the process. To this attempt, we calibrated three different groups of subjects with three different calibration images that had different visual cues but the same number of dots.

In this sense, if calibration was a process entirely based on the numerosity, as the model of Izard & Dehaene suggests, no differences between groups would be expected. Otherwise, the differences found would prove that visual cues have a relevant role in the process that need to be taken into account either as mere interferences (and so the Izard & Dehaene model should include them) or as the central element through which the numerosity process occurs (as Gebuis et al. suggest).

Hence, the objectives of the current work were:

1. Studying the effects that visual cues have in the estimation task.
2. Studying the effects that calibration has in the estimation task.
3. Studying the effects that visual cues have in the calibration.

METHODS

1. PARTICIPANTS

Sixty-three students at the University of Padova participated in the experiment after providing written informed consent. Three participants were excluded for being considered outliers, which lead to a final sample of sixty people (31 females; mean age 23 ± 2.52, range 19-32).

Participants had mixed nationalities, hence not all of them took the experiment in their mother tongue. To this extent, the groups were balanced according to the gender (F = female / M = male) and the language of the subjects (Y = experiment carried out in their mother tongue / N = experiment not carried out in their mother tongue).

GROUP 1 20		**GROUP 2** 19		**GROUP 3** 21	
GENDER					
F	**M**	**F**	**M**	**F**	**M**
9	11	11	8	11	10
LANGUAGE					
Y	**N**	**Y**	**N**	**Y**	**N**
13	6	10	9	12	9

Table 1.*Description of the sample. Number of participants included in each of the groups. Two variables were taken into account in order to balance the subjects: gender (F= female / M = male) and language of the subjects (Y= experiment carried out in their mother tongue / N = experiment not carried out in their mother tongue).*

All participants had normal or corrected-to-normal vision, and normal color vision. No participant was aware of the purpose of the experiment before taking part on it. The participants were divided into three groups and each performed one of the three conditions in a single session. The study was approved by the ethical committee of the Department of Psychology (University of Padova).

2. ALGORITHM

The stimuli were generated using a version of the program developed by de Marco, Cutini & Abalo (in prep.). We chose this program instead of the traditional ones because, as explained in the introduction, we considered that this algorithm overcomes the limitations of previous programs. Furthermore, de Marco et al. (in prep.) provides a better control of the visual cues that fitted better with the aim of our experiment, which led us to our final choice.

2.1. UNITS OF THE ALGORITHM

The units of the features that the algorithm uses are based on the Cartesian system. The Matlab function describes it parameters using a Cartesian system with axes that go from -1 to 1. Because of this, the side of the area is two units long and the total possible surface is four square units.

Hence, all the values given in the Method are expressed in these units (defined by the Cartesian system used by the algorithm). As specified below, only one of the features, *windowsize*, is instead expressed in pixels.

2.2. ARGUMENTS OF THE ALGORITHM

The Matlab function employed has six different arguments that need to be inserted as input variables. In order to better understand the visual cues control, the six arguments are briefly explained below, as well as the values give to them in order to generate our stimuli. For a more detailed description of the algorithm, see de Marco et al. (in prep.)

- **Num:** the argument *num* indicates the number of dots that compose the current image. We inserted different values according to the numerosities we wanted for our images (25, 30 and 36 for the stimuli and 18 and 50 for the fillers).

- **Windowsize:** the argument *windowsize* indicates, in pixels, the side of the square canvas in which the image is generated. Users can choose any value for the argument according to their needs but, for each manipulation, the other parameters need to be adjusted in order to respect the expected output values. For our experiment, we used a *windowsize* of 500 pixels.

- **Radiusrule:** the argument *radiusrule* admits three different values.

a) When radiusrule is set to 0: the generation of the radii is totally unconstrained. Radii are randomly generated inside of a range defined by the maximum and the minimum radius possible. Beside this, there is no other control.
b) When radiusrule is set to 1: the total surfaced covered by dots is constant across numerosities.
c) When radiusrule is set to 2: the contour length of the dots is constant across numerosities.

Given that we wanted to control the size of the dots through the contour length, we set this argument to 2.

- **Totalsurface:** the argument *totalsurface* indicates the total surface covered by dots that it is held constant when *radiusrule* is set to 1. This is an arbitrary argument, which means that users can choose any value, in relation to their needs. As we did not want to control the size of the dots through this feature, we inserted a random number for this argument, and given that the previous argument (*radiusrule*) was set to 2, the code did not use this value.
- **Totalcontour:** the argument *totalcontour* indicates the total contour length that is held constant across each numerosity when *radiusrule* is set to 2. It is also an arbitrary argument, which means that users can choose any value, in relation to their needs. As explained below, we used three different values for this feature (5, 10, 15) defining three different categories of stimuli.

- **Convexhull:** the argument *convexhull* can have two different values.

a) When convexhull is set to 0: there is no control over the position of the coordinate's center of the dots.
b) When convexhull is set to a value higher than 0: the convex hull of the output image must be held constant at that value, across each numerosity included in the set. The result of this manipulation is that every generated image have a convex hull that can vary from the input argument only inside a small tolerance range (e.g., +/- 0.025%).

Given that we wanted that our images had the smallest variation possible in terms of visual cues, we fixed the convex hull to a constant value (2).

3. STIMULI

The stimuli were arrays of white dots presented on a gray background. Each image can be classified among nine different categories that are described according to two features: *Numerosity* and *Visual Cues* (***Figure 6***). Each of these two features have three different levels, which leads to the nine different categories previously mentioned.

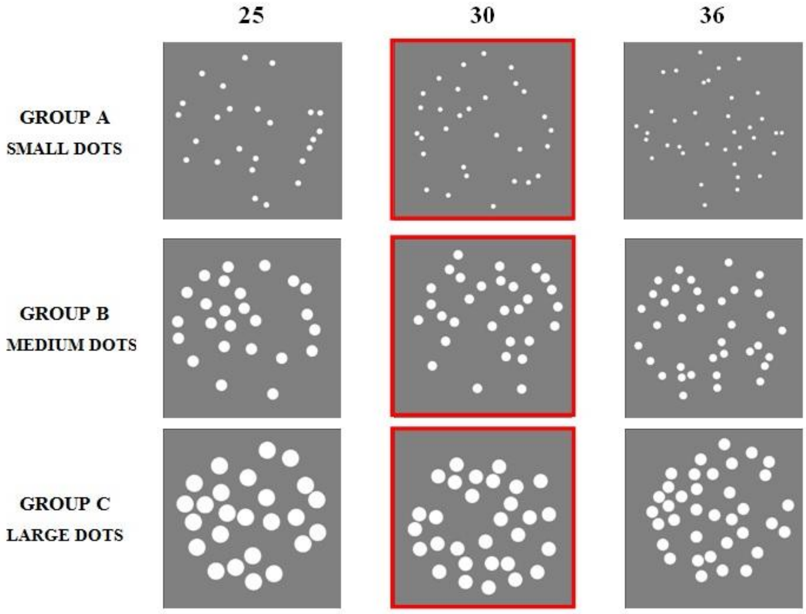


Figure 6*.Stimuli used in our set. The nine categories in which the stimuli can be classified are shown. These are formed by combining the three types of numerosity and the three types of visual cues (3x3). The category to which the calibration image belonged to is highlighted in red.*

3.1. NUMEROSITY

The *Numerosity* is the amount of dots that each image has. We selected three different numerosity values (25, 30 and 36 dots) in order to have a ratio of 5/6 between them. We decided to take this into consideration given that such a value is above the threshold that human beings are able to perceive (11/12) according to the literature (Gebuis & Van der Smagt, 2011; Halberda & Feigenson, 2008). We created categories of Numerosity different enough in order to be sure that subjects were able to perceive their differences.

Apart from that, some fillers (of 18 and 50 dots) were included into the set in order to avoid subjects to get used to the numerosities of the experiment. None of them was considered for further analyses.

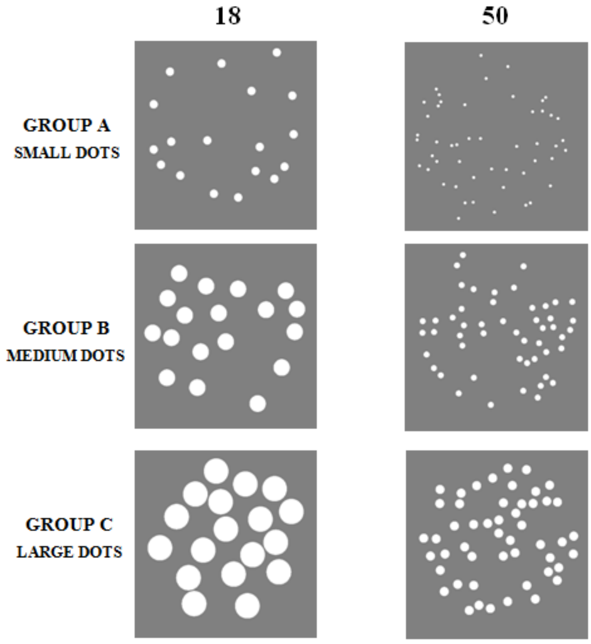


Figure 7.*Stimuli used as fillers in our set in order to avoid subjects to get used to the numerosities.*

3.2. VISUAL CUES

The stimuli were classified in three different groups according to their visual cues: Group A (or "Small dots"), Group B (or "Medium dots") and Group C (or "Large dots"). The "Size of the dots" (meaning, the *Contour Length,* the *Total Surface* and the *Diameter*) was the feature that we used in order to define these three groups, as well as the *Density* of the image. The *Convex Hull* was kept constant across the nine categories.

In other words, the three groups can be understood as:

- **Group A**: small dots with lowest Density (or highest average distance between dots).
- **Group B**: medium dots with medium Density (or medium average distance between dots).
- **Group C**: large dots with highest Density (or lowest average distance between dots).

GROUP A "SMALL DOTS"		GROUP B "MEDIUM DOTS"		GROUP C "LARGE DOTS"	
CH	2	CH	2	CH	2
DEN	0.193	DEN	0.145	DEN	0.098
CL	5	CL	10	CL	15
TS	0.067	TS	0.268	TS	0.603
DM	0.053	DM	0.107	DM	0.161

Table 2.*Visual cues data for each of the groups of stimuli. CH = Convex Hull; DEN = Density; CL = Contour Length; TS = Total Surface; DM = Diameter. The data showed for TS and DM are the average values for the three numerosities, as further explained.*

We attempted to make each group as more homogeneous as possible in terms of visual cues. This means that we tried to keep the visual cues as constant as possible even when changing numerosities. However, it is not possible to alter the numerosity of a set of dots without changing the visual cues to some extent.

For this reason, we decided to keep constant *Convex Hull* and *Density* as well as *Contour Length*, and to let *Total Surface* and *Diameter* to vary in order to alter the numerosity (*Table 2* shows in fact the average of TS and DM for the three numerosities, while the precise values for each condition are shown in *Table 3*).

In this sense and more detailed, the visual cues that defined our stimuli are:

- **Convex Hull**. *Convex Hull* was kept constant across all the numerosities and conditions (Group 1 = Group 2 = Group 3 = 2).
- **Contour Length**. *Contour Length* was kept constant across the numerosities (Group 1 = 5; Group 2 = 10; Group 3 = 15).
- **Density.** The *Density* was tried to be held constant across the numerosities but, as a consequence of the manipulation of the size of the dots, there were small differences between them (Group 1 = 0.1935 +/- 0.0035; Group 2 = 0.145 +/- 0.004; Group 3 = 0.098 +- 0.002). Even though, those differences were not significant for any group: Group 1 [$F(2,72) = 2.007$, $p = .142$], Group 2 [$F(2,72) = 0.536$, $p = .587$], Group 3 [$F(2,72) = 0.806$, $p = .216$].

– **Total surface.** *Total surface* varied across the groups and numerosities as a consequence of the previous manipulations. Their precise values are shown in *Table 3.*

– **Diameter.** *Diameter* also varied across the groups and the numerosities due to the previous manipulations. Their precise values are shown in *Table 3.*

		NUMEROSITY					
		25		30		36	
VISUAL CUES	GROUP A "SMALL DOTS"	CH	2	CH	2	CH	2
		DEN	0.193	DEN	0.193	DEN	0.193
		CL	5	CL	5	CL	5
		TS	0.08	TS	0.066	TS	0.055
		DM	0.064	DM	0.053	DM	0.044
	GROUP B "MEDIUM DOTS"	CH	2	CH	2	CH	2
		DEN	0.145	DEN	0.145	DEN	0.145
		CL	10	CL	10	CL	10
		TS	0.318	TS	0.265	TS	0.221
		DM	0.127	DM	0.106	DM	0.088
	GROUP C "LARGE DOTS"	CH	2	CH	2	CH	2
		DEN	0.098	DEN	0.098	DEN	0.098
		CL	15	CL	15	CL	15
		TS	0.716	TS	0.597	TS	0.497
		DM	0.191	DM	0.159	DM	0.133

Table 3.*Visual cues data for each of the groups of stimuli and numerosities. CH = Convex Hull; DEN = Density; CL = Contour Length; TS = Total Surface; DM = Diameter.*

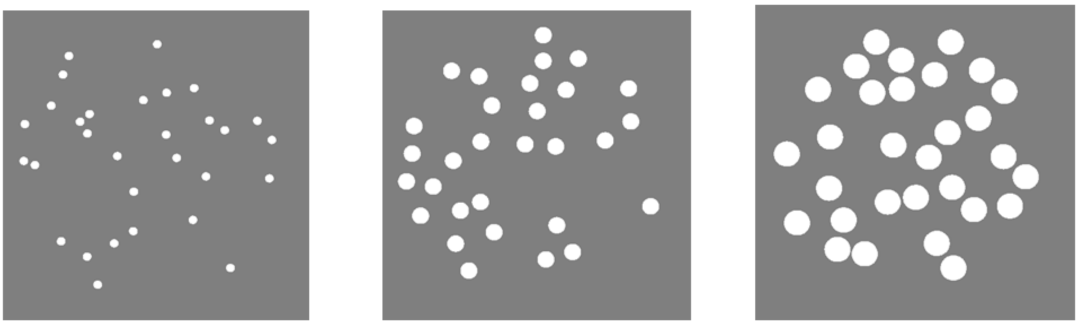
3.3. CALIBRATION IMAGE

As further explained, subjects were asked to do a calibration at the end of the first block of the experiment. The images used for such a purpose did not belong to the set of stimuli presented during the experiment, but they all shared the same features.

All of the images had 30 dots and each of them belonged to one of the three groups of visual cues described before (pointed out in red in *Figure 1*). In other words, calibration images had the same numerosities but different visual cues.

The image used to perform the calibration defined the group each participant belonged to. In this sense, and as previously mentioned, participants were divided in three different groups:

- **Group 1.** Participants calibrated with "Calibration Image 1" (30 dots - small dots).
- **Group 2.** Participants calibrated with "Calibration Image 2" (30 dots - medium dots).
- **Group 3.** Participants calibrated with "Calibration Image 3" (30 dots - large dots).



GROUP 1 GROUP 2 GROUP 3

Figure 8.*Images used for the calibration for each of the groups of participants*.

4. PROCEDURE

Participants were asked to estimate the numerosity of the arrays of dots while sitting in front of a computer. The responses of the subjects were unrestricted given that the instructions did not include any numerical value: no information related to the magnitude of the numbers involved was provided.

The experiment was constituted by two blocks: the Block without Calibration (the first one) and the Block with Calibration (the second one). The stimuli shown in each of the blocks were identical, but the order of presentation differed as the stimuli were randomized for each block. The only difference between them was that the first one was performed without any hint, while the second one was carried out after doing the calibration. In that way, participants had to estimate the numerosity of the same images, before and after calibrating.

When the "Block without Calibration" was completed, the calibration took place. After a small break, subjects were asked to look at the calibration-image for 20 seconds before facing the second block. Given that we were calibrating their responses, we informed them about the numerosity of the image (30 dots). According to the experimental group the subject belonged to, the image shown for the calibration was different: Group 1 calibrated with "Calibration Image 1", Group 2 with "Calibration Image 2" and Group 3 with "Calibration Image 3". Thus, as previously mentioned, all the participants carried out the calibration with images that had the same numerosity but different visual cues.

Each of the blocks was composed by 210 stimuli. We included 20 stimuli for each of the 9 categories previously described (3 numerosities x 3 groups of visual cues) and a total of 30 fillers was added. At the middle of each block, as well as at the end of the first one, subjects had a small break in order to avoid tiredness.

Stimuli were presented on a screen of 16 inches in a room with normal fluorescent office illumination. Participants kept a 50 cm distance from the screen. First a white fixation cross was shown for 1000 milliseconds over a gray slide. Next the array of dots was shown for 300 ms, which was followed by a gray slide with the word "Answer" written on it that remained on the screen until the answer was typed. This slide remained on the screen until the participants answered. Participants were asked to give their answers vocally to avoid any possible interference while typing them. A microphone was used to register the time of response of their answers. The numerical response was manually typed by one of the researchers.

The total duration of the experiment was approximately 45 minutes.

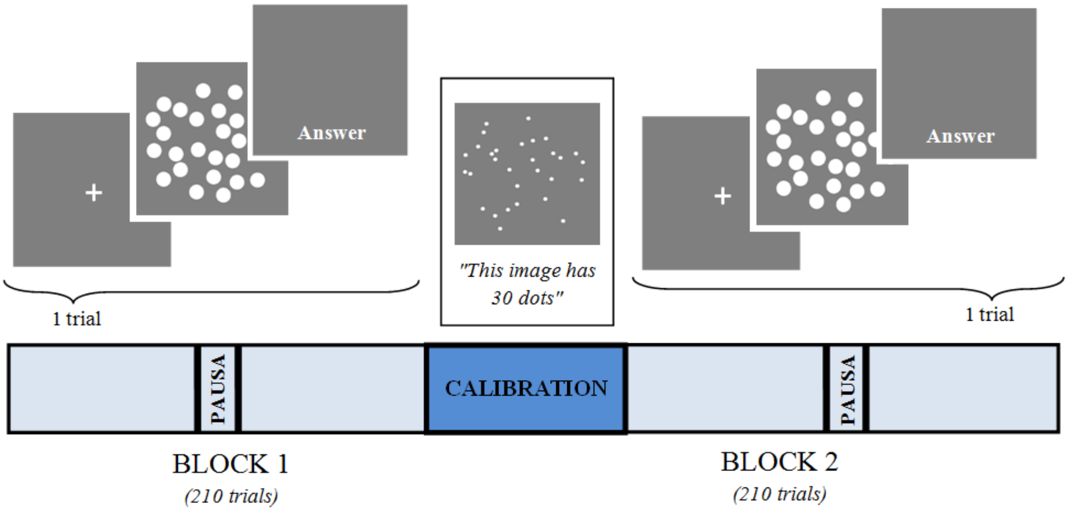


Figure 9.*Procedure of the experiment. Subjects were required to estimate the numerosities of the same 210 images before and after the calibration.*

5. <u>**ANALYSES**</u>

Delta Answers were firstly calculated for each of the trials (Delta Answer = Subject Estimate - Correct Answer). Outliers (for each participants' responses 2SD larger or smaller than the average estimate of each condition) were removed from the data as well as three subjects (one of each Group) whose answers were in more than 50% of the cases out of the confidence interval (95%). Fillers were removed from any analysis.

Then, the average response for each of the nine condition was calculated. To recall, each of the trials could be classified among nine different categories according to the features *Numerosity* (25, 30, 36) and *Visual Cues* (small dots, medium dots, large dots). Each of the nine categories included 20 different stimuli. Hence, we calculated the mean of the 20 trials for each category, so each subject had 9 different mean values.

Delta Answers means were analyzed by a mixed design Calibration (Pre-Calibration, Post-Calibration) × Group (1, 2, 3) × Numerosity (25, 30, 36) × Visual Cues (Small dots, Medium dots, Large dots) Analysis of Variance (ANOVA).

In order to understand better the interaction effects found in the previous analysis, Delta Answer were also analyzed in two new mixed design Group (1, 2, 3) × Numerosity (25, 30, 36) × Visual Cues (Small dots, Medium dots, Large dots) Analyses of Variance (ANOVAs). One of the ANOVA included the data of the Block Without Calibration and the other one the data of the Block With Calibration.

Finally, in order to understand the changes in the variability before and after the calibration, a last analysis was performed. Firstly, the *Variability in Numerosity* (VN = Delta Answer for 25 - Delta Answer for 36) for each of the three Visual Cues conditions was calculated, both before and after the calibration. Secondly, the *Variability in Visual Cues* (V VC = Delta Answer for Small Dots - Delta Answer for Large Dots) for each of the three Numerosity conditions was calculated, both before and after the calibration.

Then, four paired t-test were computed in order to understand the changes of this variability along the experiment: VN-Pre vs. Post Calibration, VVC-Pre vs. Post Calibration, VN vs. VVC Pre-Calibration and VN vs. VVC Post-Calibration. Effect size (Cohen's d) of each of these t-tests was also calculated.

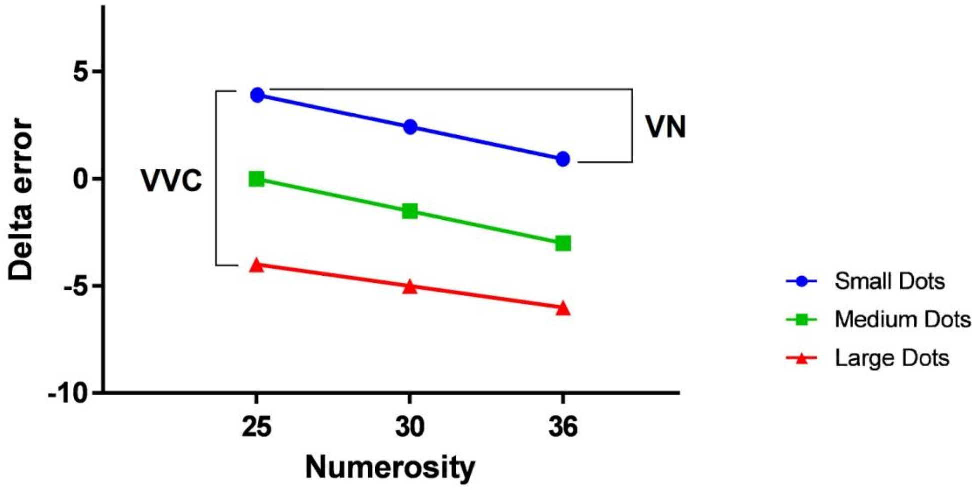


Figure 10. *Example of the two kind of variability calculated (the graph is a representation to clarify the concepts; its data does not correspond to any of the conditions). Variability in Numerosity (VN = Delta Answer for 25 - Delta Answer for 36) for the Small Dots condition is represented, as well as Variability in Visual Cues (VVC = Delta Answer for Small Dots - Delta Answer for Large Dots) for the 25 condition. Delta Answers are taken in real value and not in absolute value (e.g. VVC = 5 - (-5) = 10; VN = 4 - 1 = 3).*

RESULTS

The mixed ANOVA including all data revealed a significant main effect for the variable Calibration [$F(1,57) = 68.166$, $p<0.001$]: answers were significantly different before and after the calibration.

Furthermore, two significant interactions were found: one between Calibration and Group [$F(2,57) = 5.676$, $p = 0.006$], meaning that the mentioned Calibration effect behaved differently in the three groups, and another one between Calibration and Visual Cues [$F(2,56) = 6.867$, $p = 0.002$], meaning that the calibration affected the impact of Visual Cues on performance.

In order to have a better understanding of the interaction between Calibration and the other two variables, two separate ANOVAS were carried out, separating Pre-Calibration and Post-Calibration data. The results of these two different samples are explained in the next sections.

1. BEFORE CALIBRATION

Before the calibration, participants exhibited a general tendency of underestimating dots (named *"Underestimating Tendency"* in the discussion; see left column in Figure 11). This underestimation is larger the larger the number is (see negative slopes on left column in Figure 11).

The mixed design analysis showed a significant main effect of Numerosity [$F(2,56) =24.428$, $p<0.001$], meaning that the mistakes made for each of the categories of Numerosity differed. Furthermore, it can be seen that subjects gave a larger mean estimate for large compared to small numerosities (see positive slopes on left column in Figure 12).

A significant main effect for the variable visual cue was present [$F(2,56) =30.085$, $p<0.001$] indicating that participants gave a larger estimate for the arrays that were characterized by a relatively small size of dots (named *"Small Dots Tendency"* in the discussion; see left column in Figure 11). It is also visible that the smaller the size of the dots, the smaller the mistake is (named *"Smaller Mistake Tendency"* in the discussion; see left column in Figure11).

Importantly, we did not found an effect for the variable Group [$F(2) = 0.535$, $p = 0.588$], meaning that the three groups did not differ in a significant way before the calibration. Neither was there any interaction of this variable with the others.

2. AFTER THE CALIBRATION

Once the participants were subjected to the calibration procedure, the answers stopped being underestimated (*"Underestimating Tendency"* disappeared), being even overestimated for some of the conditions (see left column in Figure 13).

We found a significant main effect of Numerosity [$F(2,56) = 50.517$, $p < 0.001$] and Visual Cues [$F(2,56) = 18.302$, $p < 0.001$], meaning that Delta Answers continued being different between Numerosity and Visual Cues categories. Subjects continued giving a larger estimate for the arrays that had relatively small size of dots (*"Small Dots Tendency"* persisted after calibration; see right column in Figure 11). However, the smaller mistake was not present for the dots with smaller size (*"Smaller Mistake Tendency"* disappears; see Figure 11).

Furthermore, a significant effect for the variable Group was found after the calibration [$F(2) = 4.462$, $p = 0.016$]: the differences in the calibration image caused significant differences in the estimations performed by the subjects.

Finally, as expected, it can be seen that errors became smaller after the calibration, with the exception of the answers that Subjects from the Group 3 gave to the stimuli that had the smallest numerosity (see right column in Figure 13). For this particular case, subjects were more accurate when estimating the numerosity "25" before the calibration than once they have done it.

3. ANSWERS VARIABILITY

The repeated measures analyses performed with the variables VN Pre, VN Post, VVC Pre and VVC Post were all significant and all of them had a large effect size.

The *t*-tests calculated between the Variability in Numerosity Pre and Post Calibration [$t(1, 8) = -2.487$, $p = 0.038$] and the Variability in Visual Cues Pre and Post Calibration [$t(1, 8) = 5,220$, $p = 0.001$] had both a large effect size (Cohen's $d = -0.829$ and Cohen's $d = 1.740$ respectively).

The first result indicates that there was a considerable increment in the *Variability in Numerosity* after the calibration: subjects provided more different answers for the different numerosities once they had calibrated, indicating that they were more sensitive to the numerosity. Additionally, the second results shows a relevant decrement in the *Variability in Visual Cues* after the calibration. This means that subjects gave more similar answers between the different size of the dots after the calibration, indicating that they were less sensible to this variable.

The T-Tests calculated between the VN and the VVC Pre Calibration [$t(1, 8) = -3.647$, $p=0.007$] and the VN and the VVC Post Calibration [$t(1, 8) = 4.760$, $p=0.001$] had as well a large effect size (Cohen's d = -1,215 and Cohen's d = 1,586 respectively).

This results means that, before the calibration, the *Variability in Visual Cues* was larger than the *Variability in Numerosity* while, once the calibration was performed, the opposite happened. To recall, a good performance would imply a large variability in VN (given they are answers of different numerosities) and a short variability in VVC (given they are answers to the same numerosities). This is, in fact, what happens after the calibration.

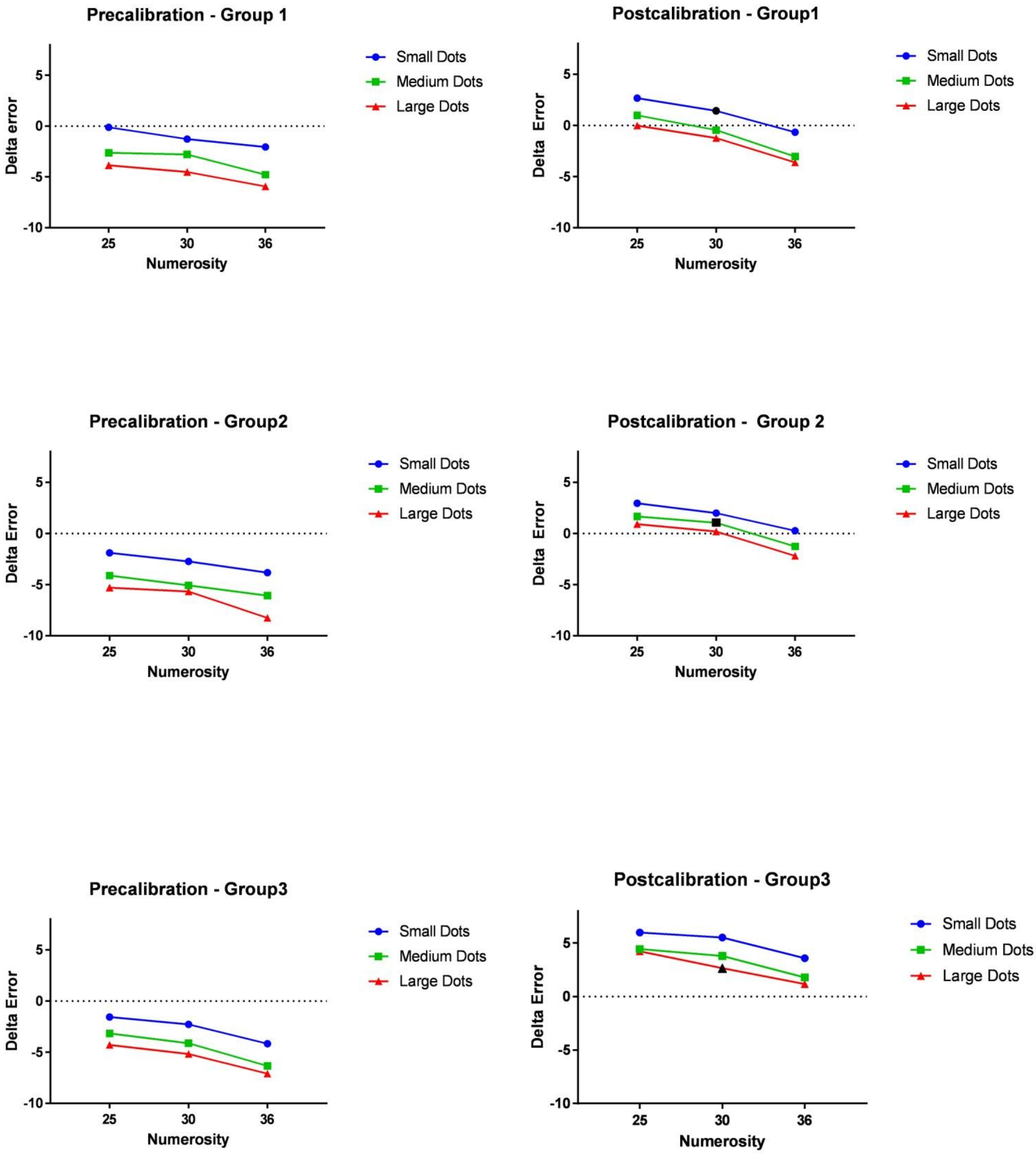


***Figure 11. Mean of delta answers for each group and condition**. Horizontal axis: numerosities; vertical axis: delta answers (delta answer = correct answer - subject's estimate); separate lines: Visual Cues groups A-B-C. The column on the left shows the answers given before the calibration and the column on the right the answers given after the calibration. The marks highlighted in black in the right column shows the category to which the calibration image belonged. The straight black dots line highlights the error 0. Above this line answers are overestimated and below it they are underestimated.*

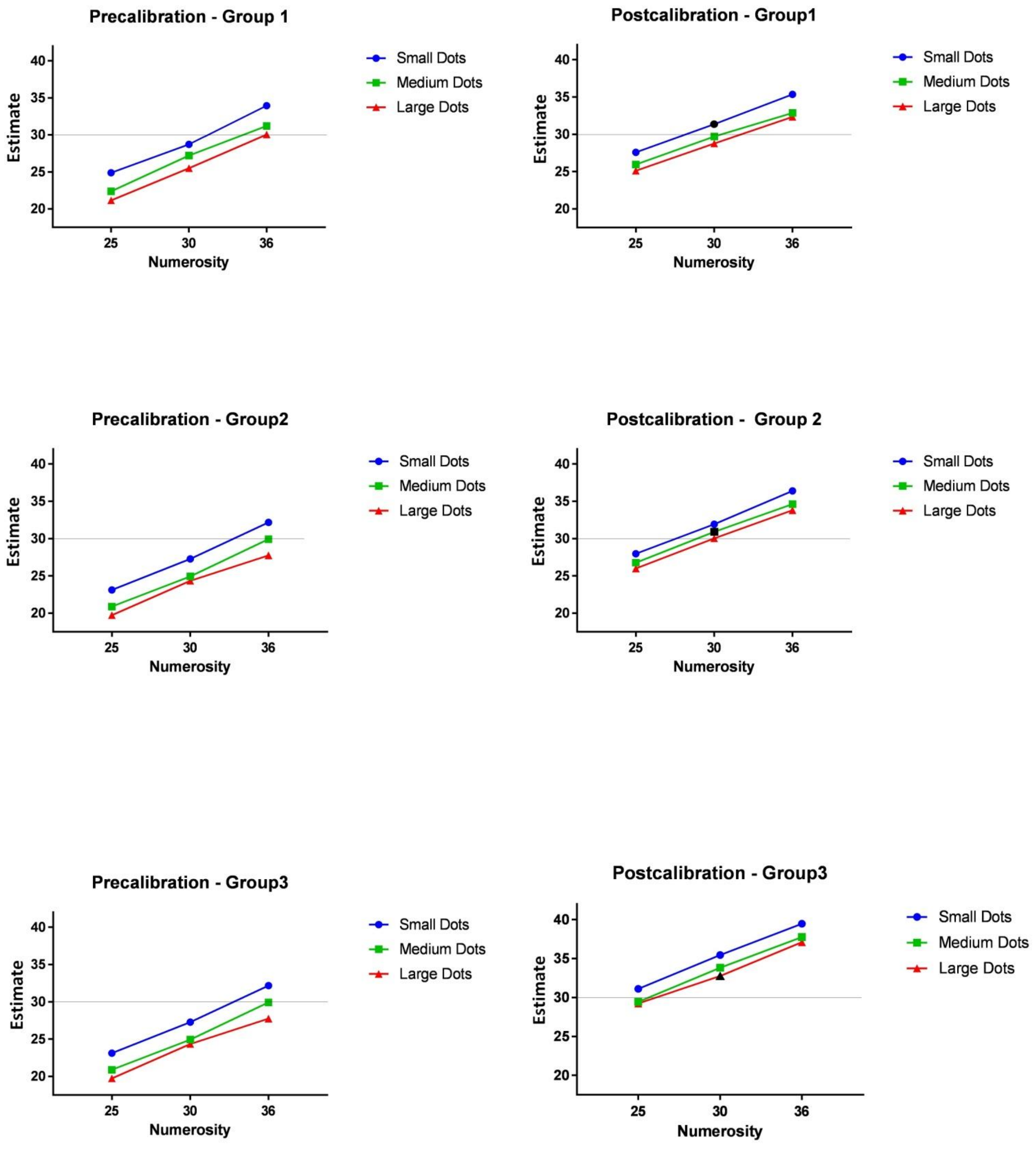


Figure 12. Mean of subjects' estimate for each group and condition*. Horizontal axis: numerosities; vertical axis: estimate; separate lines: Visual Cues groups A-B-C. The column on the left shows the answers given before the calibration and the column on the right the answers given after the calibration. The marks highlighted in black in the right column shows the category to which the calibration image belonged. The straight gray line highlights the 30 value (numerosity of the calibration image).*

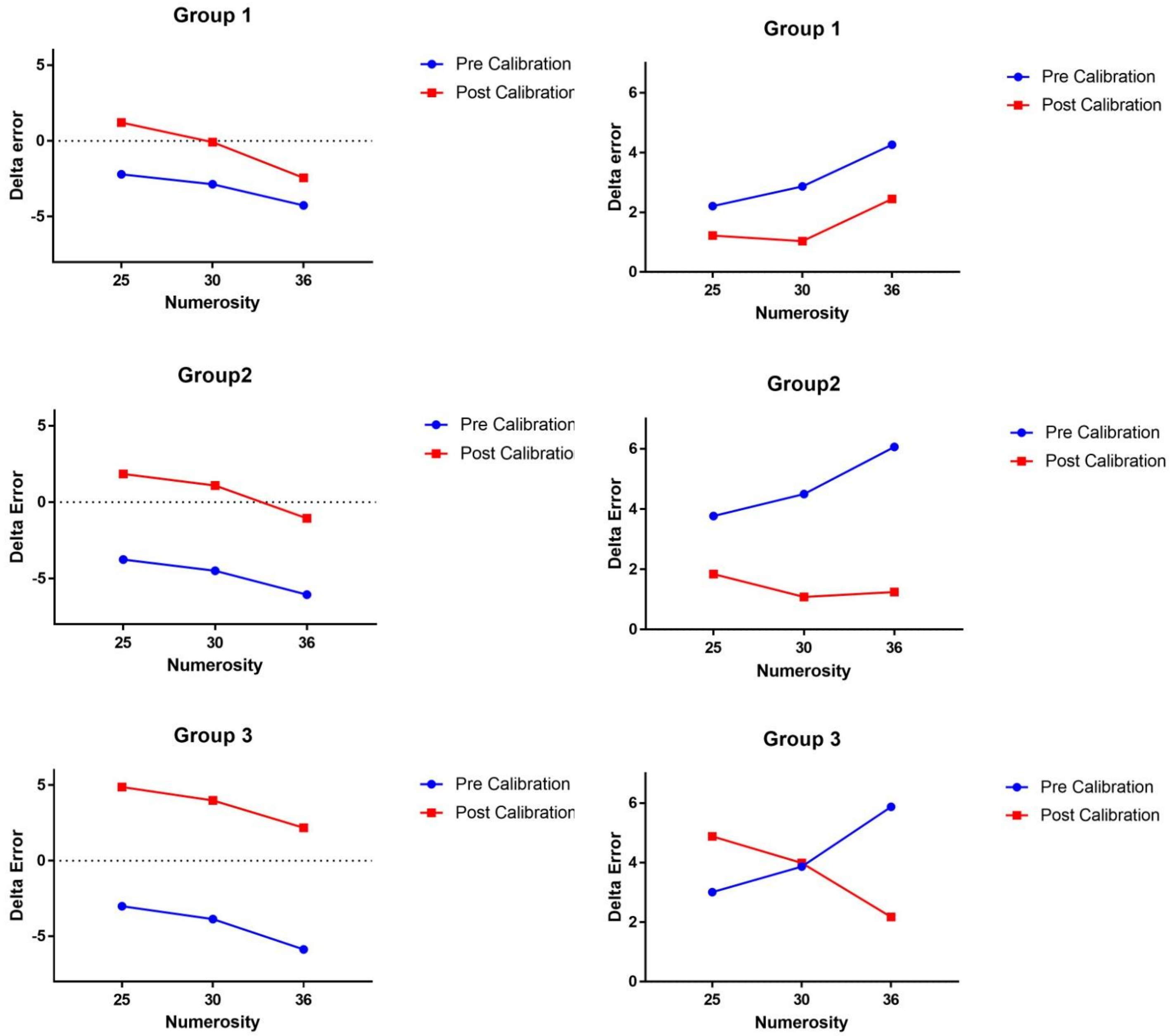


Figure 13.Calibration effects per group and condition. *Horizontal axis: numerosities; vertical axis: delta answers (delta answer = correct answer - subject's estimate); separate lines: calibration (before or after it). Each of the dots represents the mean of delta answers for all the subjects and visual cues conditions for each of the numerosities before and after the calibration. The column on the left took the delta answers values (either in positive or in negative) while the column on the right took the delta answer in absolute values.*

DISCUSSION

The importance that numerosity judgments have on daily decisions is undisputed: from doing the groceries to playing some sports, numbers are an essential piece of information that need to be processed in all of these activities. Hence, the question is: what is the nature of the processing mechanism used to arrive to such an estimate? Do we have an innate system that can extract numerosity independent of its sensory cues or rather a sensory-integration system that relies on the sensory cues present in the stimulus?

As it has been already explained, there are two current theories advocated in order to explain such phenomenon: the traditional Approximate Number System theory and the most recent Sensory-Integration theory.

According to the traditional view, we would be equipped with an ancient and innate system sensitive to large numerosities called ANS (Izard & Dehaene, 2008; Piazza et al., 2011). Even if visual cues could influence the beginning of the process to some extent, its impact would be removed later on, in a step of the process called "normalization stage". On their view, the calibration of the estimation task would increase the accuracy of the task given it provides a better mapping between the numerosity and the verbal label. Visual cues are not taken into consideration in the calibration at all (Izard & Dehaene, 2008).

However, the increasing importance of visual cues have led to a new theoretical framework that tries to understand number processing from a different perspective. According to Gebuis et. al (2016), numerosity judgments would rather rely on a sensory-integration processing which is based on the visual cues of the stimuli. In this sense, there would not be such a thing as a normalization state, given that visual cues would be always a relevant component of the process. On their view, the improvement due to calibration would be based on the visual cues: showing a reference image would provide subjects the opportunity to establish a relationship between the size of the different sensory cues and the numerosity.

In order to shed some light on this issue, in this study we investigated the role of visual cues in the calibration of an estimation task. Participants were presented with dot arrays representing 25, 30 or 36 dots and had to estimate the number of dots shown before and after they calibrated with a 30-dots image. To investigate the effects of the visual properties, we divided the subjects in three different groups. Each of the group carried out the calibration with a different image: all of them had the same numerosity of dots but different visual properties.

Our results show that the traditional theory of calibration (Izard & Dehaene, 2008) is insufficient: given that it does not consider visual cues influence, it is not able to explain our results. This does not necessarily deny the existence of an innate system able to perceive numerosity by itself, but highlights the relevance that visual cues have for the process. On the other hand, our experiment also shows that more research is necessary to explain in detail the calibration process according to the Sensory-Integration Theory. We have divided the discussion in different sections in order to improve its comprehension.

1. ESTIMATION TASKS AND VISUAL CUES

First of all, and independently from the calibration, our results replicate those found by Gebuis & Reynvoet (2012a): visual cues have an influence in estimation tasks. More precisely, they found that participants estimated that the number of dots was larger in the arrays characterized by a relatively large convex hull, small density and small total surface (or average diameter).

Our experiment confirmed those results: participants tended to overestimate the Group A of stimuli ("Small dots") compared to the other two (the already called *Small Dots Tendency*). To recall, Group A was the one characterized by a smaller total surface, a smaller average diameter and a smaller density (or bigger "average distance between dots"); the convex hull was held constant within the Group A-B-C. Hence, the direction of the bias found in our experiment was similar to previous studies (Gebuis & Reynvoet, 2012a; Gebuis & Reynvoet, 2012b).

2. ESTIMATION TASK AND CALIBRATION

Regarding the calibration, we can firstly say that we found a main effect of the variable Calibration, which shows that answers were significantly different before and after the calibration. This is in line with the theoretical view of Izard & Dehaene (2008), which highlighted the relevance that such a part has in the estimation task.

In Figure 11, the black mark highlights the category to which the calibration image belonged to. As it can be seen, there is always a slight over estimation of this category, meaning that subjects are not able to place the calibration image exactly in the numerosity "thirty", but between 1 and 3 points above it, consistently with the overestimation of 1.5 points from the calibration point found in a previous study (Izard & Dehaene, 2008).

Furthermore, calibration had some effects on the three tendencies that we described before. In first place we can notice that, once the calibration was done, *Underestimating Tendency* disappeared: instead of underestimating the arrays of dots, participants started to label the stimuli they were seeing with larger numbers. This led in some occasions to a slightly overestimation of the sets of dots.

As a consequence of this, *Smaller Mistake Tendency* also did: the smaller mistake was not anymore made in the category of smaller dots. This evidences that the tendency was not due to a higher accuracy in smaller dots but, in fact, to an artifact built for the contrast between the opposite effects that *Underestimating Tendency* and *Small Dots Tendency* have. The general underestimation of the sets of dots faced the larger estimates given to the smallest dots, which led to a smaller mistake to these last ones. Once the general underestimation had disappeared, this effect was not visible anymore, which evidences that participants were not simply more accurate with the smallest dots.

3. CALIBRATION AND VISUAL CUES

Beyond this main effect result, two interaction effects of the variable Calibration were also found (with the variable Groups and with the variable Visual Cues). The ANOVAs performed with the two levels of the variable Calibration (Pre-Calibration and Post-Calibration) showed a significant main effect of the Group with the Post-Calibration data. As explained before, groups only differed in the calibration image, thus, no difference before groups was found before the calibration. Hence, it can concluded that the differences on the visual cues of the calibration images induced differences on the estimates.

This result, together with the bigger error made for numerosity 25 by the third group of subjects after the calibration, cannot be explained by the classical theories of ANS (Izard & Dehaene, 2008). According to the traditional view, calibration improves performance in estimation tasks entirely due to numerosity factors: visual cues are not thought to have a role on the processing. For this reason, a better performance after calibration is always expected, as well as no differences caused by different visual cues if the same numerosities are kept.

In that way, our results cannot be reconciled with the classical view: it becomes mandatory to take visual cues into account during the calibration process. As already mentioned, this does not necessarily goes against the existence of an innate system able to perceive numerosity directly, but stresses the essential role that visual cues have in it. A theory that does not consider them a part of the process is not able to explain our results and, hence, it would need to be revised.

On the contrary, the results found in our study could be easily explained from a perspective that considered the relevance of visual cues, as the Sensory-Integration Theory does.

According to this view, once subjects had calibrated, they would compare the visual cues of the calibration image to the next stimuli. The answer would be given according to this comparison, so more visual cues would correspond to a larger number, and less visual cues to a smaller answer. Along the experiment, this comparison would be done either between the memory of the calibration image and the current stimuli (that could be possible given the global influence that calibration has, Izard & Dehaene, 2008) or either the previous stimuli and the current one (as Gebuis et. al. suggested).

Hence, subjects would *locate* the calibration point (highlighted in black in Figure 11) and, after it, they would "apply" the *"Small Dots Tendency"*: the smaller dots would be considered more numerous (and given respectively larger numbers) and the larger dots would be considered less numerous (and given so smaller numbers).

This is also what happened for subjects belonging to Group 3. However, in this group, the image they calibrated with belonged to the category in which the largest underestimation was made: more than five points below the correct answer (see left column of Figure 11).

As a consequence, when they calibrated, they realized about the big size of the error they were making. Given that the other categories of dots seem to be more numerous due to their visual cues (*"Small Dots Tendency"*), they started to label the sets of dots with very large numbers. In fact, this group of subjects is the only one in which thirty is understood as the lowest value: there is approximately no mean values under the number 30. On the contrary, subjects belonging to the other groups took the calibration number as an intermediate value; and, because of this reasons, there are answers above and below this number.

Hence, after applying the *"Small Dots Tendency"* in Group 3 (once the calibration point had already been settled), subjects happened to be less accurate for the numerosity 25 after the calibration: the underestimation that was made before it was, in absolute value, smaller than the overestimation made after it. In other words, the general and marked overestimation for calibration image 3 caused a largest error in the array with less dots after the calibration.

4. CALIBRATION AND VARIABILITY

In order to better understand the influence of calibration, we calculated the variability of the answers for both Numerosity and Visual Cues before and after the calibration (see section 3 of the results). The variability across numerosities and across different visual cues can be seen as a measurement of the sensitivity: the lower the variability was between the different levels of the variable the less sensitive the subject was to those changes in the variable.

For instance, a low variability in Numerosity (VN) would mean that participant's answers did not differ too much between the most extreme categories of Numerosity (that is, 25 and 36), while a low variability in the Visual Cues (VCC) would mean that answers were more or less similar between the most extreme categories of the variable (Small Dots and Large Dots). In this sense, an accurate performance would require a high VN (meaning that participants' answers highly varied between the most extreme Numerosity categories) and a low VCC (meaning that participants' answers did not substantially changed for the different size of dots).

In this line, our results show that calibration increased the sensitivity to Numerosity and decreased the bias (or big sensibility to) that Visual Cues were causing when estimating the stimuli. How is this explained by the two theoretical frameworks?

In case there was a system able to perceive the numerosity by itself, these results could be easily explained: calibration would become an indispensable part of the estimation task, given that it reduces the influence of the visual cues and improve the perception of the numerosity. Hence, subjects would have a better accuracy when estimating numerosities after calibrating.

On the other hand, in case numerosity was not something we could perceive directly and number processing would happen through the visual cues, as SIT suggests, these results, even if still explainable, would become more difficult to interpret.

According to Gebuis et al.'s view, visual cues are the basis for number processing. In that way, they would consider that the differences found in our results related to "numerosity" could be explained in fact by the differences of the visual cues that correlate with the changes in numerosity.

To recall, even if we tried to reduce the differences of visual cues inside of each group of visual features (Group A-B-C), the cues "Total Surface" and "Diameter" also changed when changing the numerosities (see Table 3 in Methods), given that it is not possible to alter numerosity while helding constant all the visual cues. Because of this reason, it is not possible to attribute completely any of the results observed to numerosity itself: the visual cues could be explaining the results as well.

In fact, from Gebuis et al. 's perspective, "numerosity" should be substituted by "visual cues that correlated with numerosity". In this sense, calibration would be changing the sensibility for the different visual cues: it would increase the sensibility for the "visual cues that correlated with numerosity" and would decrease the sensibility to the other visual cues (size of the dots).

Even though, SIT should explain with more detail how this process happens to occur. In case numerosity was not directly perceived, it would be necessary to understand which of the visual cues are actually informing us about numerosity and how they are doing it. Furthermore, it is essential to understand why the calibration influences participants' answers and though which mechanisms it does it.

5. NEURAL CORRELATES AND OUR RESULTS

Before coming to our conclusion, we would like to discuss about how our results fit with the neural correlates that have been traditionally pointed out as the biological basis of numerical processing. As previously mentioned, the intraparietal sulcus is the region thought to be the neural correlate of the ANS, both in children and adults (Piazza & Izard, 2009; Cutini et al. 2014).

Single cell recordings in monkeys corroborated these results. Nieder et al. (2002, 2003, 2004, 2007) documented neurons in the PPC and PFC of trained monkeys that were maximally responsive to their preferred number (Nieder & Miller, 2004). They also showed that their data was best described by a nonlinearly compressed scaling of numerical information, as postulated by the Weber-Fechner law or Steven's law for psychophysical/sensory magnitudes (Nieder & Miller, 2003). Furthermore, these neurons, instead of showing monotonic response functions as the ones encoding purely sensory magnitude typically do ("summation coding"), exhibited non-monotonic tuning functions that peaked at specific numerosities ("labeled-line code"). (Nieder & Merten, 2007). These results were understood as the neural correlate of the ANS.

However, later interpretations suggest that these neurons do not have to be necessarily sensitive to numerosity: they could be sensitive to visual cues instead. In this sense, the Sensory Integration Theory would be also plausible from a biological point of view.

First of all, these studies did not properly control for visual cues (Gebuis et al., 2016), so it is not possible to entirely attribute the results to numerosity. Furthermore, the parieto-frontal network for visual numerical information suggested by Nieder & Miller (2004) is similar to the visual processing network for spatial information (Thiebaut de Schotten et al., 2005).

Besides this, a later modeling study (Verguts &Fias, 2004) included a summation stage in between the sensory processing and the number selective tuning (Gebuis et al, 2016). Some years later, Chen and Verguts (2013) demonstrated that not-trained-on-numerosity-monkeys neural responses could follow a summation coding process instead of being numerosity-selective. These results make the numerosity processing to become more similar to a sensitive one. Furthermore, the Weber-Fechner law is thought to be for sensory magnitudes (like the visual cues are) more than high-level perception (like numerosity).

Despite these facts, even if neurons that specifically respond to numerosity would exist in our brain, this does not necessarily mean that they are innate (as the ANS is supposed to be): they could have been specialized within the experience. To recall, monkeys used in most of these studies were trained to perceive numerosity, so such an experience could have implied a specialization of some of their neurons.

Because of all of these reasons, we believe that neural correlates typically found for number processing are perfectly in line with our results. In that way, it does not matter if the processing is carried out by the numerosity or by the sensory properties of the stimuli: even if such a processing was guided by visual cues, this would not go against the neural correlates that are already known.

In conclusion, our results evidence the pivotal role that visual cues have in number processing. We studied for the first time the influence that visual properties have when calibrating the estimation task and so, we propose that an update of the classical views of the process that would take visual cues of calibration into consideration is highly required. Furthermore, we also showed how calibration seems to change participants' sensibility for visual cues and numerosity (or "visual cues correlated to numerosity"), increasing accuracy through decreasing the impact of the first ones and simultaneously increasing the impact of the second one. Further investigation is required in order to better understand the mechanism underneath our skills to deal with numbers.

CONCLUSIONS

- Visual cues have an essential role in number processing that has been traditionally underestimated. In estimation tasks, visual properties happen to influence not only the task but also the calibration process.

- Calibration seem to increase the low sensibility to "numerosity" and decrease the bias (or big sensibility) that visual cues were provoking when estimating the stimuli.

- It is necessary to update the classical theories that explain calibration (Izard & Dehaene, 2008) given that the process cannot only being explained exclusively through a numerical view.

- Given the increasing role that visual cues are being attributed, it is necessary to reconsider the theoretical framework of ANS. In this sense, it might be useful to consider the alternative theory suggested by Gebuis et al. (2016), even if further research is needed to support such a theory.

- Research in ANS field is suffering from severe methodological issues that need to be faced. Furthermore, the lack of ecological validity to which the control of visual cues as well as the artifacts created by so should make us reconsider the approach we are giving to the study of this cognitive skill.

- Taking into account the existent problems when doing research about ANS, it is premature to develop programs of training that try to improve Math ability through this skill. A better understanding of the numerical process is the previous and necessary step to make.

BIBLIOGRAPHY

Agrillo, C., Dadda, M., & Bisazza, A. (2007). Quantitydiscrimination in femalemosquitofish. *AnimalCognition*, 10(1), 63–70. http://dx.doi.org/10.1007/s10071-006-0036-5.

Agrillo, C., Dadda, M., Serena, G., & Bisazza, A. (2008). Do fish count? Spontaneous discrimination of quantity in female mosquitofish. *Animal Cognition*, 11(3), 495–503.http://dx.doi.org/10.1007/s10071-008-0140-9

Agrillo, C., Piffer, L., & Bisazza, A. (2010).Large number discrimination by mosquito fish. *PloS One*, 5(12), e15232.http://dx.doi.org/10.1371/journal.pone.0015232

Agrillo, C., Piffer, L., Bisazza, A., & Butterworth, B. (2012).Evidence for two numerical systems that are similar in humans and guppies. *PloSOne*, 7(2), e31923. http://dx.doi.org/10.1371/journal.pone

Agrillo, C. (2015). Numerical and arithmetic abilities in non-primate species. In R. C.Kadosh, & A. Dowker (Eds.), *The handbook of numerical cognition. Oxford University Press.*

Alvarez, G. A., & Cavanagh, P. (2004).The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological science*, *15*(2), 106-111.https://doi.org/10.1111/j.0963-7214.2004.01502006.x

Beran, M. J., Perdue, B. M., & Evans, T. A. (2015).Monkey mathematical abilities. In R. C.Kadosh, & A. Dowker (Eds.*), The handbook of numerical cognition. Oxford University Press.*

Brannon, E. M., & Roitman, J. D. (2003).Nonverbal representations of time and number in animals and human infants.*Functional and neural mechanisms of interval timing*, 143-182.

Brannon, E.M., Suanda, S., &Libertus, K. (2007). Temporal discrimination increases in precision over development and parallels the development of numerosity discrimination. *Developmental Science*, 10(6), 770–777. https://dx.doi.org/10.1111%2Fj.1467-7687.2007.00635.x

Burr, D., & Ross, J. (2008).A visual sense of number. *Current Biology*, 18(6), 425–428.http://dx.doi.org/10.1016/j.cub.2008.02.052

Burr, D., Turi, M. & Anobile, G (2010) Subitizing but not estimation of numerosity requires attentional resources. *Journal of Vision*, 10(6): 20, 1 - 10.

Butterworth, B. (2005). The development of arithmetical abilities. *Journal of Child Psychology and Psychiatry*, 46(1), 3–18. https://doi.org/10.1111/j.1469-7610.2004.00374.x

Cantlon, J. F., &Brannon, E. M. (2005).Semantic congruity affects numerical judgments similarly in monkeys and humans. *Proceedings of the National Academy of Sciences of the United States of America*, 102(45), 16507–16511. http://dx.doi.org/10.1073/pnas.0506463102

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), e125. https://doi.org/10.1371/journal.pbio.0040125

Cantlon, J. F. & Brannon, E. M., (2006) Share system for ordering small and large numbers in monkeys and humans. Psychological Science, 17(5), 401-6 https://doi.org/10.1111/j.1467-9280.2006.01719.x

Cantlon, J. F., Platt, M. L. & Brannon E. M. (2009) Beyond the number domain. *Trends In Cognitive Science*, 13 (2): 83 - 91. https://doi.org/10.1016/j.tics.2008.11.007

Cantlon, J. F., (2012) Math, monkeys and the developing brain. *ProcNatlAcadSci USA*, 109 Suppl 1:10725-32. https://doi.org/10.1073/pnas.1201893109

Clayton, S. & Gilmore, C. (2014) Inhibition in dot comparison tasks.*ZDM: the international journal on mathematics education* 47(5) DOI: 10.1007/s11858-014-0655-2

Chen, Q., & Verguts, T. (2013). Spontaneous summation or numerosity-selective coding? Frontiers in Human Neuroscience, 7, 886. http://dx.doi.org/10.3389/fnhum.2013.00886.

Cutini, S., Scatturin, P., Moro, S. B., & Zorzi, M. (2014).Are the neural correlates of subitizing and estimation dissociable? An fNIRS investigation. *Neuroimage*, *85*, 391-399. https://doi.org/10.1016/j.neuroimage.2013.08.027

Dadda, M., Piffer, L., Agrillo, C., & Bisazza, A. (2009).Spontaneous number representation in mosquito fish. *Cognition*, 112(2), 343–348. http://dx.doi.org/10.1016/j.cognition

Dakin, S. C., Tibber, M. S., Greenwood, J. A., Kingdom, F. A., & Morgan, M. J. (2012).A common visual metric for approximate number and density. *Proceedings of the National*

Academy of Sciences of the United States of America, 108(49), 19552–19557. http://dx.doi.org/10.1073/pnas.1113195108

Dehaene, S. (1997). The number sense: How the mind creates mathematics. New York, Cambridge (UK): *Oxford University Press, Penguin Press.*

Dehaene, S., Dehaene-Lambertz, G., & Cohen, L. (1998).Abstract representations of numbers in the animal and human brain.*Trends in Neurosciences*, 21(8), 355–361 (doi: S0166223698012636 [pii]).

Dehaene, S. (2003).The neural basis of the Weber-Fechner law: A logarithmic mental number line. Trends in Cognitive Sciences, 7(4), 145–147.http://dx.doi.org/10.1016/S1364-6613(03)00055-X

Dehaene, S., Izard, V., & Piazza, M. (2005).Control over non-numerical parameters in numerosity experiments 2005. *Unpublished manuscript (available on www. unicog. org)*.

Dehaene, S. (2009). Origins of mathematical intuitions: The case of arithmetic. *Annals of the New York Academy of Sciences*, 1156, 232–259.http://dx.doi.org/10.1111/j.1749-6632.2009.04469.x

De Marco, D., Cutini, S. & Abalo, I. (in prep). A new method to generate non-symbolic number stimuli.

De Smedt, B., & Gilmore, C. K. (2011).Defective number module or impaired access?Numerical magnitude processing in first graders with mathematical difficulties.*Journal of Experimental Child Psychology*, 108(2), 278–292. http://dx.doi.org/10.1016/j.jecp.2010.09.003

Emmerton, J., &Renner, J. C. (2009).Local rather than global processing of visual arrays in numerosity discrimination by pigeons (Columba livia).*Animal Cognition*, 12(3),511–526. http://dx.doi.org/10.1007/s10071-009-0212-5

Fechner, G. T. (1860). Elemente der Psychophysik. Leipzig. *Breittsopf and Hartel. I, 860.*

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in cognitive sciences*, *8*(7), 307-314.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), 74–79. http://dx.doi.org/10.1111/cdep.12019

Flombaum, J. I., Junge, J. A., & Hauser, M. D. (2005). Rhesus monkeys (Macacamulatta) spontaneously compute addition operations over large numbers. *Cognition*, 97(3), 315–325.http://dx.doi.org/10.1016/j.cognition.2004.09.004

Gebuis T., van der Smagt M. J. (2011). False approximations of the approximate number system. *PLoS ONE*, 6.http://dx.doi.org/10.1371/journal.pone.0025405

Gebuis, T., & Reynvoet, B. (2011).Generating non-symbolic number stimuli.*Behavior Research Methods*, 43(4), 981–986. http://dx.doi.org/10.3758/s13428-011-0097-5

Gebuis, T., & Reynvoet, B. (2012a). Continuous visual properties explain neural responses to non symbolic number. *Psychophysiology*, 49(11), 1649–1659. http://dx.doi.org/10.1111/j.1469-8986.2012.01461.x

Gebuis, T., & Reynvoet, B. (2012b).The interplay between non symbolic number and its continuous visual properties. *Journal of Experimental Psychology. General*, 141(4), 642–648. http://dx.doi.org/10.1037/a0026218

Gebuis, T., & Reynvoet, B. (2012c).The role of visual information in numerosity estimation. *PloS One*, 7(5), e37426.https://dx.doi.org/10.1371%2Fjournal.pone.0037426

Gebuis, T., & Reynvoet, B. (2015).Number representations and their relation with mathematical ability.In R. C. Kadosh, & A. Dowker (Eds.), *Oxford handbook of numerical cognition. Oxford University Press.*

Gebuis, T., Cohen Kadosh, R. & Gevers, W. (2016). Sensory-integration system rather than approximate number system underlies numerosity processing: A critical review. *Acta Psychologica*, 171, 17-35.http://doi.org/10.1016/j.actpsy.2016.09.003

Gilmore, C., Attridge, N., & Inglis, M. (2011). Measuring the approximate number system. *Quarterly Journal of Experimental Psychology (Colchester),* 64(11), 2099–2109.http://dx.doi.org/10.1080/17470218.2011.574710

Gilmore, C., Attridge, N., Clayton, S., Cragg, L., Johnson, S., Marlow, N., ... Inglis, M. (2013).Individual differences in inhibitory control, not non-verbal number acuity, correlate with mathematics achievement. *PloS One*, 8(6), e67374. http://dx.doi.org/10.1371/journal.pone.0067374

Gilmore, C., Attridge, N., De Smedt B. &Inglis, M. (2014) Measuring the approximate number system in children: Exploring the relationships among different tasks.*Learning and Individual Differences*, 29:50–58.DOI: 10.1016/j.lindif.2013.10.004

Ginsburg, N. (1978). Perceived numerosity, item arrangement, and expectancy.*The American journal of psychology*, 267-273.

Gómez-Laplaza LM, Gerlai R (2013) The Role of Body Surface Area in Quantity Discrimination in Angelfish (*Pterophyllumscalare*). PLoS ONE 8(12): e83880. https://doi.org/10.1371/journal.pone.0083880

Guillaume, M., Gevers, W., & Content, A. (2015). Assessing the Approximate Number System: No relation between numerical comparison and estimation tasks. *Psychological Research*. http://dx.doi.org/10.1007/s00426-015-0657-x

Halberda, J., & Feigenson, L. (2008).Developmental change in the acuity of the" Number Sense": The Approximate Number System in 3-, 4-, 5-, and 6-year-olds and adults. *Developmental psychology*, *44*(5), 1457.https://doi.org/10.1037/a0012682

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

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), 11116–11120.http://dx.doi.org/10.1073/pnas.1200196109

Halberda, J., & Ly, R. (2013). PANA math: The psychophysical assessment of number sense acuity. *Unpublished manuscript, Johns Hopkins University*.

Haist, F., Wazny, J. H., Toomarian, E., & Adamo, M. (2015).Development of brain systems for non symbolic numerosity and the relationship to formal math academic achievement. *Human brain mapping*, *36*(2), 804-826.https://doi.org/10.1002/hbm.22666

Hauser, M. D., Carey, S., & Hauser, L. B. (2000).Spontaneous number representation in semi-free-ranging rhesus monkeys. *Proceedings of the Biological Sciences*, 267(1445),829–833. http://dx.doi.org/10.1098/rspb.2000.1078

Hauser, M. D., Tsao, F., Garcia, P., &Spelke, E. S. (2003). Evolutionary foundations of number: Spontaneous representation of numerical magnitudes by cotton-top tamarins. *Proceedings of the Biological Sciences,* 270(1523), 1441–1446.https://dx.doi.org/10.1098%2Frspb.2003.2414

Inglis, M., Attridge, N., Batchelor, S., & Gilmore, C. (2011). Non-verbal number acuity correlates with symbolic mathematics achievement: But only in children. *Psychonomic Bulletin& Review*, 18(6), 1222–1229. http://dx.doi.org/10.3758/s13423-011-0154-1

Inglis, M. & Gilmore, C. (2014) Indexing the approximate number system. *Acta psychologica*, 145: 147-55. https://doi.org/10.1016/j.actpsy.2013.11.009

Izard, V., &Dehaene, S. (2008). Calibrating the mental number line.*Cognition*, 106(3), 1221–1247. http://dx.doi.org/10.1016/j.cognition.2007.06.004

Izard, V., Sann, C., Spelke, E. S., &Streri, A. (2009). Newborn infants perceive abstract numbers. *Proceedings of the National Academy of Sciences of the United States of America,*106(25), 10382–10385. http://dx.doi.org/10.1073/pnas.0812142106

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

Kaufman, E. L.; Lord, M. W.; Reese, T. W.; Volkmann, J (1949) The discrimination of visual number. *The American Journal of Psychology*, Vol. 62, 1949, 498-525.http://dx.doi.org/10.2307/1418556

Libertus, M., Feigenson, L. & Halberda, J. (2013) Is approximate number precision a stable predictor of math ability? *Learn Individual Differences*, 1 (25): 126-133. https://dx.doi.org/10.1016%2Fj.lindif.2013.02.001

Lindskog, M., Winman, A., Juslin, P. &Poom, L. (2013) Measuring acuity of the approximate number system reliably and validly: the evaluation of an adaptive test procedure. *Frontiers in Psychology*, 4:510.https://dx.doi.org/10.3389%2Ffpsyg.2013.00510

Mazzocco, M. M., Feigenson, L., &Halberda, J. (2011).Preschoolers' precision of the approximate number system predicts later school mathematics performance. *PloSOne*, 6(9), e23749. http://dx.doi.org/10.1371/journal.pone.0023749

McCloskey, M. (1992). Cognitive mechanisms in numerical processing: Evidence from acquired dyscalculia. *Cognition*, 44(1–2), 107–157.

Moyer, R. S., & Landauer, T. K. (1967). Time required for judgements of numerical inequality Nature, 215(109), 1519–1520. doi:10.1038/2151519a0

Newton, A. T., Waring, R. J., & Penner-Wilger, M. (2014). Symbiotic symbols: Symbolic (but not non-symbolic) number representation predicts calculation fluency in adults. *Austin, Texas*, 2753-8.

Nieder, A., & Miller, E. K. (2003). Coding of Cognitive Magnitude: Compressed Scaling of Numerical Information in the Primate Prefrontal Cortex. *Neuron*, 37, 149-157. https://www.ncbi.nlm.nih.gov/pubmed/12526780

Nieder, A., & Miller, E. K (2004). A parieto-frontal network for visual nuemrical information in the monkey. *Proceedings of The National Academy of Science of USA*, 11; 101(19), 7457-62. DOI: 10.1073/pnas.0402239101

Nieder, A., & Merten, K. (2007).A labeled-line code for small and large numerosities in the monkey prefrontal cortex. The Journal of Neuroscience, 27(22), 5986–5993. http://dx.doi.org/10.1523/JNEUROSCI.1056-07.2007

Park, J., & Brannon, E. M. (2013).Training the approximate number system improves math proficiency. *Psychological Science*, 24(10), 2013–2019.http://dx.doi.org/10.1177/0956797613482944

Piazza, Izard, V., Pinel, P., Le Bihan, D., & Dehaene, S. (2004). Tuning curves for approximate numerosity in the human intraparietal sulcus. Neuron, 44(3), 547–555. http://dx.doi.org/10.1016/j.neuron.2004.10.014

Piazza, & Izard, V. (2009). How humans count: Numerosity and the parietal cortex. The Neuroscientist, 15(3), 261–273. http://dx.doi.org/10.1177/1073858409333073

Piazza, Facoetti, A., Trussardi, A. N., Berteletti, I., Conte, S., Lucangeli, D., ... Zorzi,M. (2010).Developmental trajectory of number acuity reveals a severe impairment in developmental dyscalculia. Cognition, 116(1), 33–41. http://dx.doi.org/10.1016/j.cognition.

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

Piazza, M., Fumarola, A., Chinello, A. &Melcher, D. (2011). Subitizing reflects visuo-spatial object individuation capacity. *Cognition*, 121, 147 - 153.http://doi.org/10.1016/j.cognition.2011.05.007

Piazza, M., Pica, P., Izard, V., Spelke, E. S., &Dehaene, S. (2013b).Education enhances the acuity of the nonverbal approximate number system. *Psychological Science*, 24(6), 1037–1043. http://dx.doi.org/10.1177/0956797612464057

Pica, P., Lemer, C., Izard, V., &Dehaene, S. (2004).Exact and approximate arithmetic in an Amazonian indigene group. *Science*, 306(5695), 499–503. http://dx.doi.org/10.1126/science.1102085

Price, G. R., Palmer, D., Battista, C., &Ansari, D. (2012). Non symbolic numerical magnitude comparison: Reliability and validity of different task variants and outcome measures and their relationship to arithmetic achievement in adults. *ActaPsychologica*, 140(1), 50–57. http://dx.doi.org/10.1016/j.actpsy.2012.02.008

Reivking, S. K., Piazza, M., Izard V., Cohen, L &Dehaene, S. (2008) Does subitizing reflect Numerical estimation? *Psychological Science*, 19 (6): 607-14. https://doi.org/10.1111/j.1467-9280.2008.02130.x

Roggeman, C., Santens, S., Fias, W., & Verguts, T. (2011). Stages of non symbolic number processing in occipitoparietal cortex disentangled by fMRI adaptation. The Journal of Neuroscience, 31(19), 7168–7173. http://dx.doi.org/10.1523/JNEUROSCI.4503-10.

Ross, J. (2003).Visual discrimination of number without counting.*Perception*, *32*(7), 867-870.

Ross-Sheehy, S., Oakes, L. M., & Luck, S. J. (2003). The development of visual short term memory capacity in infants.*Child development*, *74*(6), 1807-1822.

Rousselle, L., & Noel, 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), 361–395. https://doi.org/10.1016/j.cognition.2006.01.005

Rugani, R., Vallortigara, G., &Regolin, L. (2014). From small to large: Numerical discrimination by young domestic chicks (*Gallus Gallus*). *Journal of Comparative Psychology*,128(2), 163–171. http://dx.doi.org/10.1037/a0034513

Santens, S., Roggeman, C., Fias, W., & Verguts, T. (2010). Number processing pathways in human parietal cortex.*Cerebral Cortex*, *20*(1), 77-88. https://doi.org/10.1093/cercor/bhp080

Sasanguie, D., De Smedt, B., Defever, E., & Reynvoet, B. (2012a). Association between basic numerical abilities and mathematics achievement.*British Journal of Developmental Psychology*.

Scholl, B. J. (2001). Objects and attention: The state of the art. *Cognition*, *80*(1), 1-46

Smets, K., Gebuis, T., Defever, E., & Reynvoet, B. (2014). Concurrent validity of approximate number sense tasks in adults and children. *Acta Psychologica*, 150, 120–128.http://dx.doi.org/10.1016/j.actpsy.2014.05.001

Smets, K., Sasanguie, D., Szücs, D., & Reynvoet, B. (2015). The effect of different methods to construct non-symbolic stimuli in numerosity estimation and comparison. *Journal of Cognitive Psychology*, *27*(3), 310-325. DOI: 10.1080/20445911.2014.996568

Spelke, E. S., & Kinzler, K. D. (2007).Core knowledge.*Developmental science*, *10*(1), 89-96.

Starr, A. B., Libertus, M. E., & Brannon, E. M. (2013a). Infants show ratio-dependent number discrimination regardless of set size. *Infancy*, 18(6).http://dx.doi.org/10.1111/infa.12008

Starr, A., Libertus, M. E., & Brannon, E. M. (2013b).Number sense in infancy predicts mathematical abilities in childhood. *Proceedings of the National Academy of Sciences of the United States of America*.http://dx.doi.org/10.1073/pnas.1302751110

Szűcs, D., Soltesz, F., Jarmi, E., &Csepe, V. (2007). The speed of magnitude processing and executive functions in controlled and automatic number comparison in children: Anelectro-encephalography study. *Behavioral and Brain Functions*, 3, 23.

Szűcs, D., & Soltesz, F. (2008).The interaction of task-relevant and task-irrelevant stimulusfeatures in the number/size congruency paradigm: An ERP study. *Brain Research*, 1190, 143–158.

Szűcs, D., Nobes, A., Devine, A., Gabriel, F. C., & Gebuis, T. (2013). Visual stimulus parameters seriously compromise the measurement of approximate number system acuity and comparative effects between adults and children. *Frontiers in psychology*, 4, 444.https://doi.org/10.3389/fpsyg.2013.00444

Thiebaut de Schotten, M., Urbanski, M., Duffau H., Volle, E., Lèvy R., Dubbois, B. & Bartolomeo, P. (2005) Direct Evidence for a Parietal-Frontal Pathway Subserving Spatial Awareness in Humans. *Science*, 309 (5744), 2226-2228. DOI: 10.1126/science.1116251

Todd, J. J., & Marois, R. (2004).Capacity limit of visual short-term memory in human posterior parietal cortex. *Nature*, *428*(6984), 751-754. https://doi.org/10.1038/nature02466

Verguts, T., & Fias,W. (2004). Representation of number in animals and humans: A neural model. Journal of Cognitive Neuroscience, 16(9), 1493–1504. http://dx.doi.org/10.1162/0898929042568497.

Vonk, J., &Beran, M. J. (2012). Bears "count" too: Quantity estimation and comparison in black bears (Ursus Americanus). *Animal Behaviour*, 84(1), 231–238. http://dx.doi.org/10.1016/j.anbehav.2012.05.001

Wood, J. N., & Spelke, E. S. (2005). Infants' enumeration of actions: Numerical discriminationand its signature limits. *Developmental Science*, 8(2), 173–181.

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

Xu, F., Spelke, E. S., & Goddard, S. (2005). Number sense in human infants.Developmental Science, 8(1), 88–101.

Xu, Y., & Chun, M. M. (2006).Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature*, *440*(7080), 91-95. https://doi.org/10.1038/nature04262