Running Head: NASCO METHOD TO GENERATE DOT ARRAY

# Title: NASCO: A new method and program to generate dot arrays for non-symbolic number comparison tasks.

Authors: Mathieu Guillaume, Christine Schiltz, Amandine Van Rinsveld.

Manuscript Word Count: 6,708 (excluding abstract, footnotes, and references).

Number of Figures: 3. Number of Tables: 3.

Keywords: Methodology, Approximate Number System, Numerical comparison task, Number Sense, Software solution.

## Abstract

Basic numerical abilities are generally assumed to influence more complex cognitive processes involving numbers, such as mathematics. Yet measuring non-symbolic number abilities remains challenging due to the intrinsic correlation between numerical and non-numerical dimensions of any visual scene. Several methods have been developed to generate non-symbolic stimuli controlling for the latter aspects but they tend to be difficult to replicate or implement. In this study, we describe the NASCO method, which is an extension to the method popularized by Dehaene, Izard, & Piazza (2005). Their procedure originally controlled for two visual dimensions that are mediated by Number: Total Area and Item Size (i.e., $N = TA/IS$). Here, we additionally propose to control for another twofold dimension related to the array extent, which is also mediated by Number: Convex Hull Area and Mean Occupancy (i.e., $N = CH/MO$). We illustrate the NASCO method with a MATLAB app – NASCO app – that allows easy generation of dot arrays for a visually controlled assessment of non-symbolic numerical abilities. Results from a numerical comparison task revealed that the introduction of this twofold dimension manipulation substantially affected young adults' performance. In particular, we did not replicate the relation between non-symbolic number abilities and arithmetic skills. Our findings open the debate about the reliability of previous results that did not take into account visual features related to the array extent. We then discuss the strengths of NASCO method to assess numerical ability, as well as the benefits of its straightforward implementation in NASCO app for researchers.

Abstract Word Count: 249.

# NASCO: A new method and program to generate dot arrays for non-symbolic number comparison tasks

In 1997, Dehaene formulated that humans possess a *Number Sense* – a biologically determined ability allowing the representation and the manipulation of large numerical quantities. Most authors currently consider that such numerical intuition relies on a cognitive system specifically dedicated to number processing (following Feigenson, Dehaene, and Spelke, 2004). This view is supported by extensive empirical evidence showing that humans can discriminate numerical quantities from early age (Xu & Spelke, 2000), with limited knowledge of number words (Pica, Lemer, Izard, & Dehaene, 2004), or without formal instruction (Nys, Ventura, Fernandes, Querido, Leybaert, & Content, 2013). Recent studies further supported this perspective by showing that humans have a spontaneous preference for the numerical aspect of large sets rather than for other continuous visual features (Cicchini, Anobile, & Burr, 2016; Ferrigno, Jara-Ettinger, Piantadosi, & Cantlon, 2017). Notwithstanding such findings, some authors challenged the existence of a specific cognitive system devoted to numerical processing and alternatively ventured that *Number Sense* emerges from the combined weighting of continuous perceptual dimensions available in visually displayed stimulus collections (Gebuis, Cohen Kadosh, & Gevers, 2016; Leibovich, Katzin, Harel, & Henik, 2016).

The debate about the *Number Sense* nature is still ongoing (see Núñez, 2017, for an interesting view) because there is a peculiar methodological issue relative to non-symbolic number comparison tasks: it is empirically impossible to isolate the cognitive processes specifically dedicated to numerical discrimination from those related to other continuous magnitude discrimination. The numerosity (*i.e.*, the information about the number of elements) is indeed intrinsically intertwined with non-numerical magnitudes, such as the luminance or the extent of the array (see for instance, Gebuis and Reynvoet, 2012). Previous studies showed that numerical judgments are substantially impacted by the total surface occupied by all items (Guillaume, Nys, Mussolin, and Content, 2013), by the individual size of the elements (Henik, Gliksman, Kallai, & Leibovich, 2017), by the item density (Dakin, Tibber, Greenwood, Kingdom, and Morgan, 2011), and by the size of the convex hull (*i.e.*, the smallest convex polygon encompassing all elements; Norris, Clayton, Gilmore, Inglis, and Castronovo, 2018).

Critically, empirical data showed that the procedure used to handle the correlation issue between magnitudes considerably influences participants' judgments, and subsequently the measurement of approximate numerical ability (Smets, Gebuis, Defever, & Reynvoet, 2014; Smets, Sasanguie, Szücs, & Reynvoet, 2015). A recent meta-analysis confirmed that the measure of participants' precision during numerical comparison tasks is tightly related to the generation algorithm used to create dot arrays (Guillaume & Van Rinsveld, 2018). This is worrying since the reliable but moderate association observed between approximate numerical discrimination and math ability (Schneider *et al.*, 2016) might be drastically affected by the way non-symbolic stimulus sets are created (Norris & Castronovo, 2016; see also, Clayton, Inglis, and Gilmore, 2018).

**Designing non-symbolic number stimulus: NASCO method.**

Piazza, Izard, Pinel, Le Bihan, and Dehaene (2004) were among the first authors to design a non-symbolic number comparison task tackling this methodological issue. Their publication was shortly followed by an unpublished document, which we will refer to as Dehaene, Izard, and Piazza (2005)[1]. This document describes how the authors manipulated two critical perceptual dimensions that are intrinsically related to Number (N): the Individual Size (*IS*) and the Total occupied Area (*TA*). It is noteworthy that in their method, the size of every item within an array was homogeneous (*i.e.*, all geometrical forms had the exact same perimeter, area, and circumference). It follows that the total area covered by the dots – in pixels – is exactly equal to the number of pixels within each dot times the number of dots. One could rewrite the previous expression as follows: Number is equal to the total area occupied by the dots divided by their individual size of the dots, or $N = TA/IS$.

The main property of the relation between TA and IS is its proportionality: for a given N, if we reduce IS, TA decreases. This natural relation between TA and IS becomes problematic when one wishes to specifically manipulate N, which is the case in research on numerical abilities. In particular, any change on numerosity

[1] Most published articles following the methodological document from Dehaene *et al.* (2005) used dot arrays, although similar construction rules were described for other basic geometrical forms, such as squares. For the sake of brevity, we focus on constructing dot arrays because we used dot arrays in the current study.

systematically impacts one of the continuous dimensions, since $a \times N = a \times (TA/IS)$. To take this relation into account, Dehaene and colleagues proposed to keep one dimension constant while letting the other freely vary. For instance, doubling N implies either to double TA and to keep IS constant [$2N = 2TA/IS$], or alternatively to divide IS by two and to keep TA constant [$2N = TA/(IS/2)$]. Dehaene and colleagues suggested keeping one dimension constant for half of the items, and the other dimension constant for the other half, so that participants could not reliably respond based on these dimensions. Participants would indeed perform at chance level if they systematically responded following a given single dimension.

In the original document from Dehaene *et al.* (2005), the authors only took into consideration the above-mentioned relation, $N = TA/IS$. Here we complement the initial approach by suggesting to consider in a similar manner another relation between two perceptual dimensions intrinsically mediated by Number: Convex Hull area (CH) and Mean Occupancy[2] (MO), considering $N = CH/MO$. In the continuity of the work by Dehaene and colleagues, one can control for these dimensions either by keeping CH constant and letting MO vary with N, or alternatively by keeping MO constant and letting CH vary with N. Critically, both CH and MO are independent from TA and IS, considering the physical constraint

---

[2] Mean Occupancy is actually the inverse of Density (*i.e.*, 1/D), the number of items within a given space. MO described as the following relation $MO = CH/N$ should be understood as the mean space that each dot sustains within and around its physical size (see Allik and Tuulmets, 1991).

that CH needs to be greater than TA (and MO needs to be greater than IS) to avoid overlapping dots. In other words, for any given values of N, TA, and IS, we can theoretically construct arrays with infinite different values of CH and MO, as long as CH > TA and MO > IS. For instance, we could draw ten 10px dots, occupying in total 100px, close together within a convex hull of 500px for a mean occupancy of 50px, or alternatively we could draw the same dots within a larger space, such as a convex hull of 5000px for a mean occupancy of 500px. By crossing the relation between TA and IS to the relation between CH and MO, it is possible to manipulate and objectively measure the relative contributions of these four major visual properties to non-symbolic number comparison decisions, with *N = TA/IS = CH/MO.* Figure 1 illustrates how we can define four categories built from the crossing of two relations mediated by Number. One can thus generate four categories of dot arrays, each for a quarter of the stimulus set.

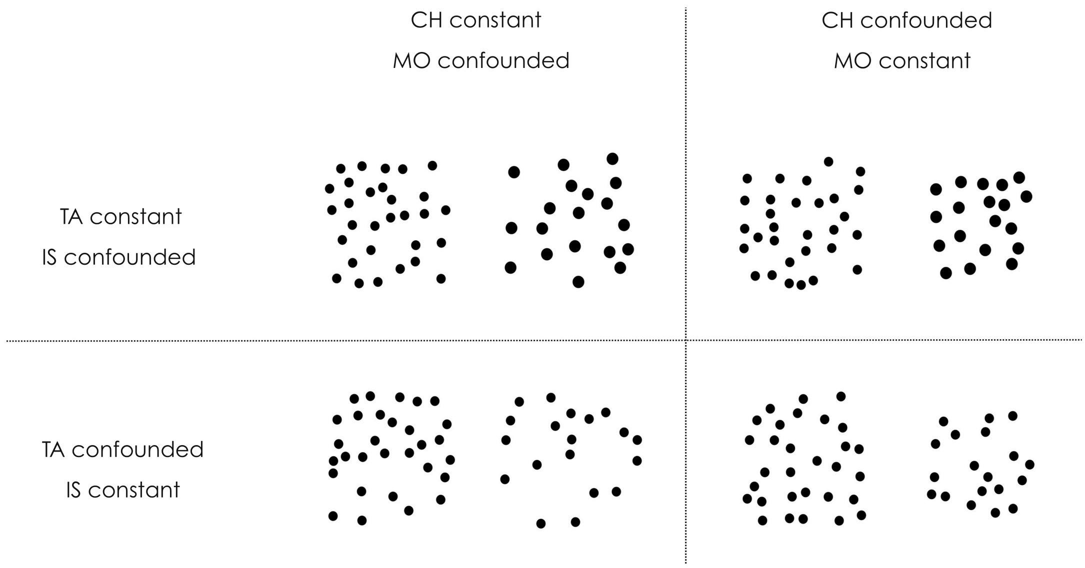


**Figure 1. Each dot array pair belongs to one of the four categories, depending on the manipulation of two bi-dimensional factors. As in Dehaene et al. (2005), either Total Area (TA) or Item Size (IA) is kept constant between both arrays of a pair; the other dimension**

**varies with numerical change. Additionally, either Convex Hull (CH) or Mean Occupancy (MO) is kept constant between both arrays of a pair; the other dimension varies with numerical change. All array pairs of the figure contain 30 vs. 19 dots.**

We name our dot generation method NASCO, as it controls for Number, Area, Size, Convex hull, and Occupancy. In the continuity of existing methods, NASCO aims at manipulating the unwanted visual dimensions in a more satisfying way than the original procedure suggested by Dehaene and colleagues (2005). Amongst these methods, one could refer to the study by Holloway and Ansari (2009) who manipulated the density of the array (but not CH) in addition to TA and IS; or to the study by Mussolin, Nys, Leybaert, and Content (2012) who displayed collections of elements of a more complex nature in order to heterogeneously vary IS. More recently, Salti and colleagues (2017) also took into account CH, and the density (*i.e.*, 1/MO), but they viewed both as being indices of the same extrinsic dimension (*i.e.*, the extent). Consequently, the method from Salti and colleagues does not allow specifically manipulating CH or MO, while our method does because it considers them separately.

Another elegant procedure to disentangle numerosity from other non-numerical dimensions is the modelling approach of DeWind and colleagues (2015). The authors considered the same relation between the dimensions as we do: they grouped TA and IS within a dimension called “Size”, and they categorized CH and MO (called sparsity in their study) within a dimension called “Spacing”. They proposed to create stimulus sets across which the numerical dimension is kept orthogonal with (*i.e.*, independent from) “Size” and “Spacing”. In other words, the authors decided to model composite dimensions resulting from the combination of the related visual dimensions (on one hand, IS and TA; on the other hand, CH and MO), whereas in NASCO method we choose to explicitly emphasize the proportional nature of their relation in an ecological manner. We made this decision because "Size" and "Spacing" dimensions do not actually refer to any real percept (which is in line with the absence of brain responses to any of these mathematically constructed dimensions reported by Park, 2018). Furthermore, this sophisticated procedure is difficult to implement since it should be joined with an adapted computation of the Weber Fraction that disentangles the numerosity contribution from the respective contributions of the two orthogonal dimensions. It is additionally difficult to replicate since the authors did not provide any script to generate dot arrays with this method.

**Creating non-symbolic number stimulus with NASCO app.**

In the previous section, we discussed the methods used to design non-symbolic number stimulus sets. In this section, we focus on software solutions allowing the creation of dot arrays (*i.e.*, the generation and presentation of image files). We will not further develop how the script from Dehaene and colleagues (2005) works since we previously provided its rationale in details. We rather describe and discuss two well-designed generation algorithms.

First of all, we need to mention *Panamath* (Halberda, Ly, Wilmer, Naiman, & Germine, 2012), which is one of the most (if not the most) commonly used programs in the literature (see Guillaume & Van Rinsveld, 2018). *Panamath* is a ready-to-use non-symbolic number comparison task, which generates dot arrays at the beginning of each recording session. A first practical limitation is that

*Panamath* does not allow exporting dot arrays outside the recording session, so it cannot be used to generate image files. Secondly, *Panamath* generates by default[3] dot arrays in a very similar way than Dehaene and colleague's method: half of the trials have (on average) the same Item size, while the other half have (on average) the same Total Area. We specify "on average" as *Panamath* differs from the original Dehaene's script in a way that dot sizes are heterogeneous within an array: *Panamath* indeed allows a random variation of each individual Item Size of maximum 20% of the mean size. Critically, since *Panamath* follows Dehaene and colleagues's (2005) method, it has the same limitations (see Clayton et al., 2018; Dakin et al., 2011; Norris & Castronovo, 2016; Norris et al., 2018). In other words, *Panamath* does not control for the array extent and the density within the image (*i.e*, CH and MO). There is no such parameter that the user can access to manipulate these dimensions.

The generation algorithm of Gebuis and Reynvoet (2011) on the other hand considered similar visual dimensions as we do in the current study: TA, IS, CH, and density (*i.e.*, 1/MO). It should be noted that their program generates stimuli where dots have different sizes within an array, so that the authors also distinguish mean circumference from mean diameter, while both are confounded in our design since all dots have the same size within a given array. The main

[3] Note that we consider here the default parameters, although the user is able to modify some properties (such as the difficulty level, the dot colour, and the presentation time). Nevertheless, the user cannot precisely specify the extent of the Convex Hull or the Mean Occupancy.

characteristic of Gebuis and Reynvoet's algorithm is that it automatically generates pairs of dot arrays in which each of the controlled dimensions is congruent with the number of elements for half of the set, and incongruent for the other half. In other words, for non-symbolic number comparison tasks, the program creates blocks of pairs where the more numerous array occupies the larger surface in 50% of the cases, but where it occupies the smaller surface in 50% of the cases (and similarly for the other dimensions under consideration). Across dimensions, a stimulus can be fully or partially (in)congruent. It is noteworthy that this process is completely automatized across the stimulus set; the user cannot manually set the nature of the relation between the visual cues for a given pair. Further, the script only considers non-numerical dimensions to be either congruent or incongruent, and the user cannot specify to which degree a given array should be more or less (in)congruent. This point is an important limitation since non-numerical ratio effects on the numerical judgement have been reported for area (*e.g.*, Guillaume et al., 2013; Nys & Content, 2012) and convex hull (Gilmore, Cragg, Hogan, & Inglis, 2016).

NASCO app aims at overcoming the latter issue by highlighting the intrinsic relation between the visual dimensions in a simple way, so that the user can easily create stimulus sets. It is important to note that NASCO app emphasizes the proportional relation between all dimensions: since $N = TA/IS = CH / MO$, *then* $a \times N = a \times (TA/IS) = a \times (CH/MO)$. As a function of the user preference, doubling N will either double TA or divide IS by two, and will either double CH or divide MO by two. Within a stimulus pair, the numerical ratio is thus systematically equal to the ratio of the changing continuous dimensions. This implies that the weight of all continuous dimensions changes is necessarily and objectively the same as the weight of the numerical changes.

NASCO app has three functionalities: First, in the single array creation mode, and thanks to the straightforward relation between all considered numerical and non-numerical dimensions ($N = TA/IS = CH/MO$), the user can specify the value of each dimension, and NASCO directly shows how any change on a given dimension affects other dimensions. To create a stimulus set, the user can (for instance) specify for each array the desired Number of dots (N), their Individual Size (IS), and their Convex Hull (CH)[4]. The values of the other dimensions – in this case, Total Area (TA) and Mean Occupancy (MO) – will automatically be computed based on the introduced values. For instance, if the user wants to generate arrays with one hundred 10px dots within an area of 100,000px, they just need to introduce N = 100, IS = 10, and CH = 100,000; NASCO will automatically illustrate that TA = 1000, and MO = 10,000 in this case. The second mode allows the user to generate pairs of dot arrays: here they only need

[4] It should be noted that NASCO app generates random positions during each iteration. In practice, it is thus not reliable to generate a given value of the size of CH with the precision of one pixel; an approximate value is rather provided (it is theoretically possible to get a CH with the wanted size but it may require a lot of iterations). By default, we tolerate a maximal error of only 1% for determining the size of CH (*e.g.*, a desired CH of 100,000px will actually range from 99,000px to 101,000px). The users can modify the value of the maximal error at their best convenience.

to specify the characteristics of one array (i.e., Number and other visual dimensions) and NASCO will automatically fit the properties of the second array as a function of the desired control parameters (either TA or IS constant, and either CH or MO constant). This functionality is illustrated in Figure 2. Finally, the third functionality allows automatized generation and display processes, in which the user only needs to enter the wanted numerical quantities. NASCO will then automatically generate the stimulus set and will be ready for displaying them and recording participants' responses. Both the generation code (created with MATLAB, The MathWorks) and the NASCO user interface are freely available at https://osf.io/axmw2/.

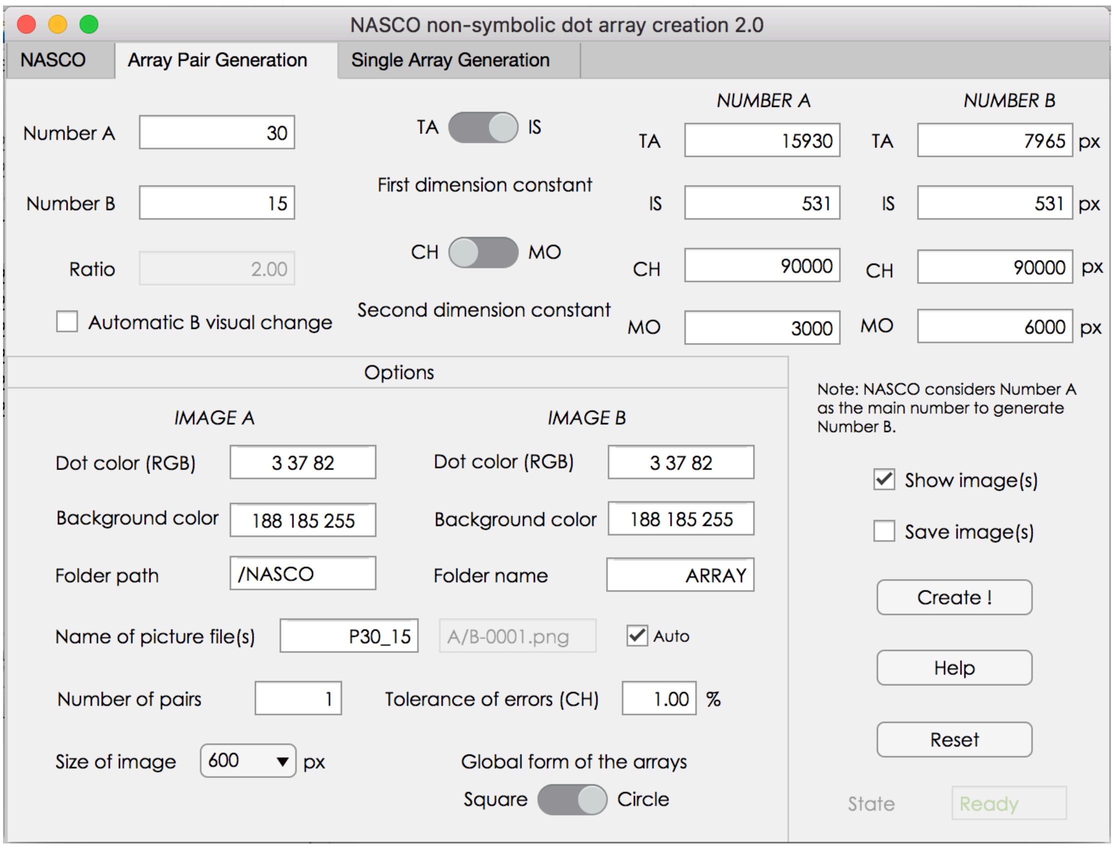


**Figure 2. Interface of the *Array Pair Generation* tab from the NASCO app. The interfaces of the other two tabs are available in Supplementary Material.**

In the current study, we illustrate the use of NASCO app by generating dot arrays with it as described in the previous section. We want to emphasize that the use of NASCO app is not limited to the generation of dots that follows the NASCO method. Since the user can set the four visual properties at its own discretion, it is possible to generate dot arrays following other recommendations. For instance, NASCO app can generate arrays following DeWind and colleagues (2015) recommendations, or alternatively can create congruent and incongruent trials as in Gebuis & Reynvoet (2012). Regarding the last possibility, since NASCO app emphasizes the relation between the visual dimensions, the user can specifically define to which degree each trial is (in)congruent , which overcomes the limitations of Gebuis & Reynvoet's original script.

**Empirical evaluation of NASCO method: Objectives and hypotheses**

We conducted an empirical study to assess the NASCO method on actual participants. The objectives of this study were twofold. Firstly, we aimed at providing a methodological evaluation of the non-symbolic stimuli designed by NASCO method and generated with NASCO app. We used these stimuli in a numerosity judgment task where participants were instructed to respond to the most numerous dot array. Since we aimed at assessing the approximate numerical ability of the participants, we expected to observe a numerical ratio effect (*i.e.*, increasing performances with increasing numerical ratio between the two magnitudes under consideration). More critically, the additional manipulation of CH and MO in our stimuli compared to previous research allowed us to directly verify whether these dimensions affected behaviour. In line with Gilmore, Cragg, Hogan, and Inglis, (2016), we expected substantial influences on numerical judgment. If this were the case such results would challenge the conclusions of previous studies that did not control for these additional visual dimensions.

Secondly, we aimed at identifying the domain-general cognitive abilities related to numerical comparison tasks under investigation. Some authors indeed surmised that the different procedures to generate dot arrays in numerical comparison tasks involve different cognitive processes, such as inhibitory control (Clayton & Gilmore, 2014). Somewhat related, inhibitory control was shown to correlate with math achievement (Gilmore *et al.*, 2013). To shed further light on this issue, participants underwent a variety of cognitive tasks assessing abilities reported to be closely related to mathematical skills: arithmetic problem solving, symbolic number processing, and executive functions (Archambeau & Gevers, 2018; Stevenson, Bergwerff, Heiser, and Resing, 2014). In this exploratory approach, we aimed at assessing whether comparison performances using our adapted stimuli were specifically related to math ability and symbolic numerical cognition, or alternatively related to domain-general cognitive abilities.

# Method

## Ethical considerations

We followed APA ethical standards to conduct the present study. The Ethic Review Panel from the Université Libre de Bruxelles approved the methodology and the implementation of the experiment before the start of data collection.

## Participants

Seventy-two undergraduate students participated in exchange of course credits (58 women, mean age was 20.36 years). Participants did not report any uncorrected visual impairment or any math disability (or history of math learning disability). In our analyses, we had to exclude one participant who failed responding to the inhibition task due to severe misunderstanding of the instructions (she systematically responded to the *no-go* trials while never responding to the *go* trials), for a final sample of seventy-one participants.

## Apparatus

Participants were tested in a large room in groups of five to six people, for an approximate duration of forty-five minutes. Each participant sat in front of a computer screen, isolated from the other ones with the help of separation panels. All tasks except the paper-and-pencil arithmetic test were displayed on a computer screen with MATLAB (The MathWorks), using the Psychophysics Toolbox extension (Brainard, 1997; Kleiner, Brainard, Pelli, Ingling, Murray, Broussard, & 2007; Pelli, 1997). All participants started with the arithmetic test, and then took part in the computer tasks, whose order was randomized across participants. Each computer task started with several trials with feedback as examples, which were not comprised in the analyses. Stimuli were displayed on a 19-in screen with a pixel resolution of 1280 × 1024px. Responses were recorded through an ioLab Systems button box. All statistical analyses were conducted with the *lme4* package (Bates, Maechler, Bolker, & Walker, 2015) for R (R Core Team, 2016).

**Arithmetic test**

We assessed arithmetic fluency with the Tempo-Test Rekenen (TTR, De Vos, 1992). This timed paper-and-pencil arithmetic test consists in five columns of forty arithmetic problems. The item difficulty increases throughout the test, from single-digit arithmetic facts such as 2 + 1 to more complex two-digit problems such as 54 + 27. The five columns of the TTR encompass one column per operation (addition, subtraction, multiplication, and division) and a final column mixing all operations. For each column, participants are instructed to write down as many correct responses as they can in one minute. Participants are awarded one point per correct answer. The maximum score of this test is 200.

**Non-symbolic stimuli and experimental task**

We specifically generated non-symbolic stimulus pairs by using NASCO app (see Introduction). We generated 192 dot array pairs divided in four stimulus categories of 48 pairs each, see Figure 1. We took arrays of 30 dots as the standard numerosity to which the second array was compared. We created the second arrays by computing six numerical ratios (from 1.1 to 1.6 with an incremental step of 0.1) starting from the standard numerosity in both increasing and decreasing directions. Crucially, by design, there were thus six non-numerical ratios, since the ratios of the changing continuous dimensions were equal to the numerical ratios. The number of dots ranged from 19 to 48, and there were 32 pairs for each ratio (*i.e.*, sixteen where the other numerosity was below 30, and sixteen where it was above 30). All dots had the same size within an array. Across the stimulus set, mean IS was 547px, Range (R) [348, 860px]; mean TA was 16420px, R [10239, 25975px]; mean CH was 112276px, R [69674, 178389px]; and mean MO was 3746px, R [2287, 5895px]. The position of the more numerous dot array of the pair (*i.e.*, the correct response) was randomly assigned to the left or to the right throughout the experiment.

We presented pairs of dot arrays and participants were instructed to determine as accurately as possible the array that contained the greater number of dots, by pressing the button on the side of the larger quantity. The onset of each trial was preceded by a fixation cross appearing 500ms before the dots. Although speed was not emphasized, the dot arrays only remained on the screen for a maximal duration of 800ms; they were then suppressed by an active mask displayed until participant's response. The mask was followed by a blank screen for 400ms, for an inter-stimulus interval of 900ms (including the fixation cross). We analysed both accuracies and response times, but we only considered Correct Response rates (CR) for correlation analyses since they sufficiently depicted the performance at this task. We did not compute the Weber fractions, as recent evidence suggested they are not more informative than accuracies (Inglis & Gilmore, 2014; Guillaume & Van Rinsveld, 2018).

## Symbolic comparison

We assessed symbolic number processing with a number symbol comparison task similar to the one by Holloway and Ansari (2009). Participants had to compare seventy-two pairs of single-digit numbers ranging from 1 to 9. Both digits were simultaneously displayed on both sides of the screen. Participants were instructed to press the button corresponding to the side of the larger digit as quickly and accurately as possible. The numerical distance within digits of the pairs ranged from one to six, resulting in twelve pairs per distance. We considered the Inverse Efficiency Score (*IES*, Townsend & Ashby, 1978) in our analyses to consider both accuracies and response times. We computed individual *IES* by dividing the mean response time of each participant by his/her mean correct response proportion.

## General processing speed

We evaluated general processing speed with a match-to-sample task (for a similar task see Hoffman, Mussolin, Martin, & Schiltz, 2014). Participants were instructed to rapidly compare one central target shape (either a circle or a diamond) to two possible solution shapes simultaneously displayed at the left and at the right of the screen. They had to identify as quickly as possible the solution shape that was identical to the target by pressing the leftmost or the rightmost button of the response box. We considered average Response Times (RT) to the correct trials as the general processing speed.

## Visuo-spatial working memory

We assessed visuo-spatial Working Memory (VSWM) because of their well-documented link to the acquisition of number skills (Cornu, Schiltz, Martin, & Hornung, 2018; Geary, 2011). We adapted a paradigm based on the no-grid task by Martin, Houssemand, Schiltz, Burnod, and Alexandre (2008). In this task, participants were instructed to remember the spatial locations of black dots briefly and sequentially displayed on a 4 × 4 invisible grid (sixteen possible locations). After each dot sequence, a fixed configuration consisting of the same number of dots was displayed. Participants had to evaluate whether the given configuration was identical to the spatial locations of the dots previously presented. Half of these configurations corresponded to the preceding sequence; the other half differed in the location of one dot from the sequence. Participants were asked to press the leftmost button if the given configuration was identical to previous series, or the rightmost button if otherwise. Critically, the number of dots to be memorized – and thus the WM load – progressively increased throughout the task, from 3 to 6 dots within one sequence. There were 36 trials in total. In the correlation analyses, we computed the sensitivity index by subtracting the False Alarm rate (FA) from the Hit Rate (HR) to have an individual measure of the visuo-spatial WM ($d' = Z(\text{HR}) - Z(\text{FA})$, Macmillan & Creelman, 2005).

**Inhibition task**

To assess inhibitory control, we adapted the task of Georges, Hoffmann, and Schiltz (2016). This task involves inhibition processes at two different levels because participants perform a Stroop-like judgement (Stroop, 1935) following Go/No-go instructions. More specifically, there were experimental and catch trials: On experimental trials, a coloured horizontal arrow pointing either to the left or to the right was presented. On catch trials, a coloured diamond was displayed for two seconds before the start of the next trial. There were 60 experimental trials, and 16 catch ones. Participants were instructed to respond to the colour of the arrow irrespective of its direction and to refrain from responding to the diamond. Critically, the buttons matching the colour of the shape, red and blue, were respectively on the leftmost and on the rightmost side of the response box. The irrelevant spatial dimension (*i.e.*, the direction of the arrow) was congruent with the response laterality for half of the trials, and incongruent for the other half. In the correlation analyses, we computed the Inverse Efficiency Score (*IES*, Townsend & Ashby, 1978) by dividing the congruent and incongruent response times by their corresponding proportion accuracies. Finally, to get one inhibition measure per participant, we calculated *IES* differences between congruent and incongruent trials (Δ *IES*). A greater Δ *IES* reflected worse performance on the latter than on the former, showing lower inhibition performances.

# Results

## Control tasks

Descriptive statistics for all control tasks are summarized in Table 1. The paper-and-pencil arithmetic test only produced a raw score for each participant as it was timed. Table 1 further reports general accuracy and mean correct RT for the computerized control tasks, and the additional measures computed for the symbolic digit comparison task, the VSWM task, and the Inhibition task (IES, *d*', and Δ IES).

**Table 1. Descriptive statistics for control tasks**

| Task | Measure | *M* | *SD* | 95% *CI* |
|---|---|---|---|---|
| Symbolic comparison | Accuracy | .965 | .183 | [.960, .970] |
| | Correct RT | 0.453 | 0.155 | [0.449, 0.458] |
| | IES | 0.468 | 0.064 | [0.452, 0.483] |
| Arithmetic test | Raw score (out of 200) | 128 | 23 | [122, 133] |
| General processing speed | Accuracy | .956 | .204 | [.945, .966] |
| | Correct RT | 0.527 | 0.137 | [0.520, 0.535] |
| Visuo-spatial WM | Accuracy | .763 | .425 | [.746, .779] |
| | Correct RT | 1.903 | 1.426 | [1s.840, 1.966] |
| | *d'* | 1.675 | 0.944 | [1.451, 1.898] |
| Inhibition | Accuracy | .960 | .194 | [.955, .965] |
| | Correct RT | 0.649 | 0.340 | [0.640, 0.658] |
| | Δ IES | 66.478 | 67.424 | [50.519, 82.437] |

***Note*: Accuracies are depicted in proportion from 0 to 1; Correct Response Times are expressed in second. *M* = Mean, *SD* = Standard Deviation, *CI* = Confidence Intervals**

## Non-symbolic numerical magnitude judgments

Overall, participants correctly detected the more numerous array of dots in 89% of the cases, 95% *CI* [88.6, 89.6], with an average latency of 657 milliseconds, 95% *CI* [650, 664]. As expected, the numerical ratio affected performance; for the smallest ratio (i.e., 1.1), performances dropped to a mean accuracy of 77%, 95% *CI* [.753, .787], and mean correct RT increased to 754 milliseconds, 95% *CI* [739, 770]. Conversely, the largest ratio (1.6) led to the best performance, with a mean accuracy of 97%, 95% *CI* [96.3, 97.7] and mean correct RT at 602 milliseconds, 95% *CI* [592, 613]. More relevant to the purpose of the current study, the stimulus properties significantly affected performance. The effects of the experimental manipulations are depicted on Table 2. Participants performed the non-symbolic magnitude judgments better when Total Area (TA) and Convex Hull (CH) were confounded (i.e., not constant) with number.

**Table 2. Behavioural performances (i.e., accuracy and correct RT) during the non-symbolic comparison task as a function of the stimulus dimensions (Total Area/Dot Size in one hand, Convex Hull/Mean Occupancy in the other hand) that were confounded (i.e., not constant) with number**

| First Dimension Confounded | Second Dimension Confounded | Accuracy | Correct RT |
|---|---|---|---|
| Total Area | Convex Hull | .944 [.937, .952] | 618 [604, 632] |
| Total Area | Mean Occupancy | .894 [.884, .904] | 647 [637, 658] |
| Dot Size | Convex Hull | .913 [.903, .922] | 669 [648, 689] |
| Dot Size | Mean Occupancy | .813 [.800, .826] | 701 [689, 712] |

***Note*: Accuracies are depicted in proportion from 0 to 1; Correct Response Times are expressed in millisecond. Brackets indicate 95% CI.**

We analysed the statistical effects of both numerical ratio and stimulus properties with linear mixed effect models. We constructed two full models (*i.e.,* one for accuracy, one for latency) with both numerical ratio and stimulus properties as fixed effects (without interactive form), and with participants as random factor and random intercept. We used logistic regression to model accuracy. We inspected the residual plots for latency models to ascertain that there were no obvious deviations from homoscedasticity or normality. To assess the significance of each main factor, we compared the full models to two reduced models without the factor in question using chi-square tests on the log-likelihood values. The full models fitted significantly better than the models without the numerical ratio factor, $\chi^2$ (5) = 758.06, $p < .001$, and $\chi^2$ (5) = 208.06, $p < .001$, for accuracy and latency respectively; meaning that the ratio significantly impacted performances. Stimulus properties also had a significant effect, as the full models were significantly better than the reduced ones, $\chi^2$ (5) = 374.07, $p < .001$, and $\chi^2$ (9) = 103.26, $p < .001$. Finally, we assessed interactions between the two main factors by comparing both full models with and without interactive form. The interaction was significant for accuracy, $\chi^2$ (4) = 31.92, $p < .001$, see Figure 3, but it was not for correct RT, $\chi^2$ (4) = 4.132, $p = .388$.

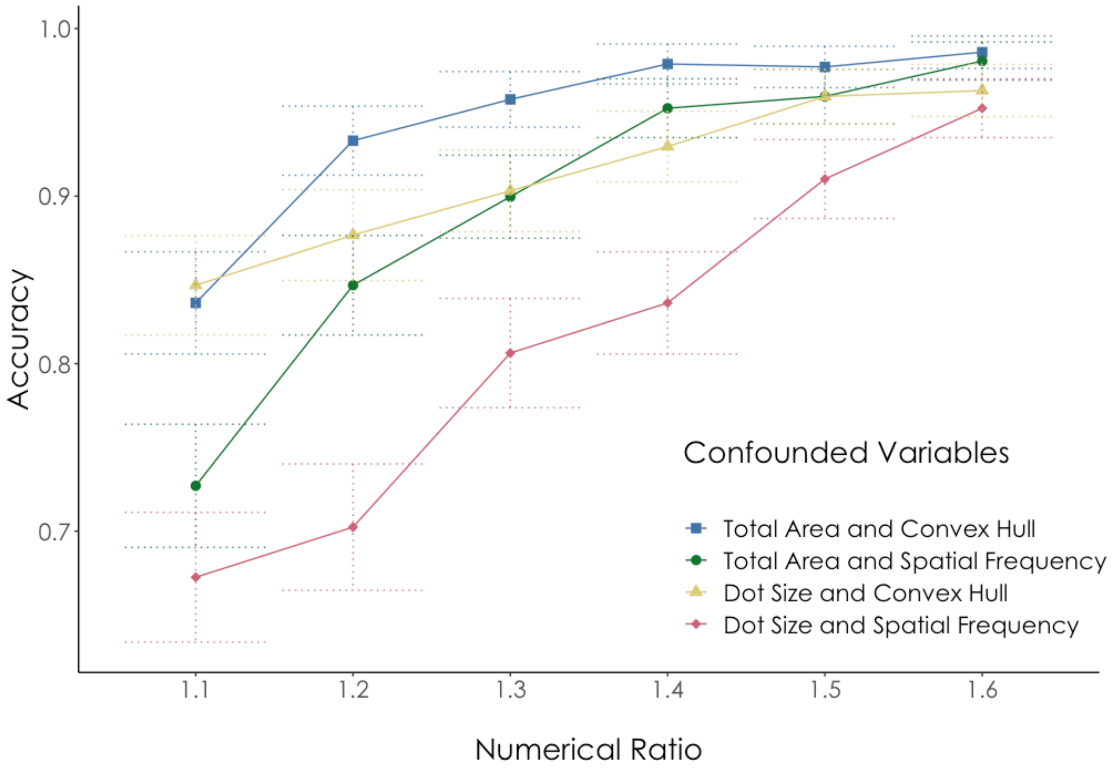

**Figure 3. Mean accuracy (in proportion) as a function of the numerical ratio and depending on the visual dimensions that were confounded (i.e., not constant) with number. Dotted vertical lines depict 95% CI.**

Overall, participants performed significantly better when Total Area and Convex Hull were confounded with number. We further looked at performance across all trials to disentangle the impact of these two cues. On one hand, we grouped all trials where TA varied with numerosity (irrespective of CH/MO, lower part of Figure 1), and on the other hand, we grouped all trials where CH varied with numerosity (irrespective of TA/IS, right part of Figure 1). We found that participants correctly responded to items in which TA was confounded with number in .916, 95% CI [.913, 926] of the cases, in 632ms, 95% CI [623, 641], whereas they performed at .928, 95% CI [.922, 935], in 643ms, 95% CI [631, 655] for trials in which CH was confounded. The comparison of confidence intervals reveals that accuracies (but not latencies) were significantly different between the conditions, which supports that Convex Hull had a more beneficial effect than Total Area. This finding is in line with previous results that Convex Hull has a stronger impact than total area on numerical judgments (Gilmore et al., 2016).

## Correlations analyses

We considered one measure per task to conduct correlation analyses: mean Correct Response rates (CR) of the non-symbolic magnitude judgments, raw scores of the arithmetic test, and latencies of the general processing speed task. For the other tasks, we computed other measures that combined response times and accuracies (*IES*, in symbolic digit comparison task), one that considers task specificities (d’ in the VSWM task), or both (Δ IES in the inhibitory control task). We focused on Kendall’s τ correlation coefficient between the variables, as it was shown to be robust to outliers (Croux & Dehon, 2010). Table 3 summarizes the coefficients.

**Table 3. Kendall’s τ correlation coefficients between performance in the non-symbolic number comparison task and performance in the other control tasks (N = 71).**

| | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|
| (1) Non-symbolic number comparison (CR) | -.032 | .007 | -.081 | .247* | -.169* |
| (2) Symbolic digit comparison (IES) | | -.148 | .289* | -.202* | .101 |
| (3) Arithmetic (Raw Score) | | | -.198* | .177* | -.025 |
| (4) General processing speed (Correct RT) | | | | -.207* | .064 |
| (5) Visuo-spatial working memory (*d’*) | | | | | .001 |
| (6) Inhibitory control (Δ IES) | | | | | |

**Note. An asterisk depicts statistical significance at the bilateral threshold of .05. Except for measures involving IES (*i.e.*, (2) and (6)), a greater value is associated with better performance.**

Correlational analyses revealed that performances of non-symbolic magnitude judgments did not correlate significantly with arithmetic performances, τ = .007, N = 71, *p* = .928, nor with symbolic magnitude judgments, τ = -.032, *p* = .690, nor with general processing speed, τ = -.081, *p* = .320. We nonetheless found significant correlations between non-symbolic magnitude judgments’ accuracy and both visuo-spatial working memory, τ = .247, *p* = .003, and inhibition, τ = -.169, *p* = .039. More generally, it should be noted that visuo-spatial working memory significantly correlated with most of our measures (see Table 3).

## Discussion

In this study, we designed a new non-symbolic stimulus generation method – NASCO – extending the recommendations from Dehaene et al. (2005) by taking into consideration visual parameters that were not included in the original document, *i.e*. the extent of the convex hull and the mean occupancy. Using a non-symbolic stimulus set specifically designed with NASCO app in a numerical magnitude judgment task with young adults, we replicated the well-known numerical ratio effect on performance: closer numerical magnitudes were more difficult to compare than more distant ones. The replication of the numerical ratio effect across the trials – even while controlling for additional visual dimensions - suggests that participants indeed performed numerical judgments during the task and therefore supports the validity of our stimulus generation algorithm.

Moreover, manipulation of the Individual Size (IS) and the Total Area (TA) influenced the numerical magnitude judgments. Participant performed very well when TA varied together with numerical magnitude (*i.e.*, when IS was kept constant across the patches). Conversely, performances dropped when IS varied with number (*i.e.*, when this time TA was kept constant). These observations are line with previous reports that TA is a visual dimension that significantly affects numerical magnitude judgments (*e.g.*, Gebuis and Reynvoet, 2012). They also support Gebuis and colleagues' (2016) criticism that averaging data from the half of items where one dimension is manipulated with data from the other half where the other dimension is controlled is insufficient to set aside the alternative hypothesis that numerical judgments are based on one of the manipulated visual cue (see also Leibovich et al., 2016). The fact that TA was the stronger dimension (in comparison to IS) in our design is not surprising since TA was confounded with the luminance of the array, which is a very salient feature in visual perception (Krauskopf, 1980), whereas IS was previously found not to influence performance above the subliminal threshold (Gilmore et al., 2016). In addition to that, our stimulus set ranged from 19 to 48 dots, and some authors reported that density (and therefore MO) has a stronger influence when the number of elements is much larger (hundreds of dots, see for instance Dakin et al., 2011). One critical remaining question is whether the influence of TA/IS is automatic and implicit, or rather strategic and task-driven. A recent study emphasized that participants deliberately and strategically use the non-numerical visual dimensions to make their numerical judgment, which is even more worrying for the reliability of the non-symbolic comparison task (Roquet & Lemaire, 2019). This issue should be further investigated in future studies.

More critically for the purpose of the current study, which proposes to consider and control also the Convex Hull (CH) and the Mean Occupancy (MO) of dot sets, the manipulation of the latter two visual attributes was not negligible. It substantially influenced numerical comparison performances. Participants consistently had more difficulties in judging numerical magnitude when CH was kept constant (*i.e.*, when MO varied). Alternatively, one could say that participants were better to compare numerical magnitudes when CH was confounded with number (*i.e.*, when MO was kept constant). In other words, participants responded as a function of the extent of the array, which follows the natural law “more items take more place”. This result is in line with previous reports that convex hull might be an even more influential cue than IS or TA (Gilmore et al., 2016). In our dataset, CH was indeed more impactful than TA, which strengthens the necessity to control for this aspect when designing dot arrays (see also Clayton & Gilmore, 2015).

As we clearly observed, the manipulation of two additional visual dimensions to the classic method of Dehaene et al. (2005) drastically affected performances. This has important implications for the literature as many studies used the original method or *Panamath,* and they thus did not take into account these influential visual dimensions, which were randomly varied throughout the experiment. If we take the hypothetical situation of designing a study where only TA and IS are manipulated, then the impact of CH and MO on behaviour would be missed. Our findings thus corroborate the concern from some authors (Gebuis et al., 2016; Leibovich et al., 2016) that we might need to critically reconsider many previously published results. This concern is even more pressing regarding the results of the current study in terms of correlation analyses. With the present dataset, we were not able to replicate any correlation between our measure of approximate numerical ability (that comprises TA, IS, CH, and MO) and math ability or symbolic magnitude judgments, which should be moderately related according to a meta-analysis (Schneider et al., 2016). However, we found a significant correlation between our measure of non-symbolic numerical ability and domain-general abilities such as visuo-spatial working memory and inhibitory control. This finding is consistent with the criticism that these processes are implied during non-symbolic numerical comparison tasks (Inglis & Gilmore, 2011), and supports Norris and colleagues'

(2018) concern that the measurement of numerical ability in the literature might be too biased to be informative. Nevertheless, the systematic correlation analysis between all these factors was not our primary objective, therefore future studies will need to investigate this issue more in details.

As a final reminder note, NASCO method does not aim at isolating the numerical dimension from every other visual dimension, or at suppressing the influence of the latter. NASCO method and app were designed for researchers or practitioners who want an easy and straightforward way to generate dot arrays. We suggest them to use NASCO app to create stimulus set that follows NASCO method. Researchers in need of sophisticated control method could still use a more elaborate method such as for instance the one from DeWind and colleagues (2015). Fortunately, NASCO app was also designed to allow such researchers to easily generate stimulus set according to their needs. We hope that this new design method and the generation algorithm will provide future guidance in designing cleaner stimulus sets, and subsequently will improve the quality and the validity of non-symbolic numerical magnitude judgment tasks.

## Conflict of Interest Statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

# References

Allik, J., & Tuulmets, T. (1991). Occupancy model of perceived numerosity. *Attention, Perception, & Psychophysics*, *49*(4), 303–314.

Archambeau, K., & Gevers, W. (2018). (How) Are Executive Functions Actually Related to Arithmetic Abilities? In A. Henik and W. Fias (Eds.), *Heterogeneity of Function in Numerical Cognition* (p. 337-357). Acadamic Press: Oxford, UK.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*(4), 433-436.

Cicchini, G. M., Anobile, G., & Burr, D. C. (2016). Spontaneous perception of numerosity in humans. *Nature Communications*, *7*, 12536.

Clayton, S., & Gilmore, C. (2014). Inhibition in dot comparison tasks. *ZDM*, 47(5), 759-770.

Clayton, S., Inglis, M., & Gilmore, C. (2018). Developmental differences in approaches to nonsymbolic comparison tasks. *Quarterly Journal of Experimental Psychology*, 1747021818755296.

Cornu, V., Schiltz, C., Martin, R., & Hornung, C. (2018). Visuo-spatial abilities are key for young children's verbal number skills. *Journal of Experimental Child Psychology*, *166*, 604-620.

Croux, C., & Dehon, C. (2010). Influence functions of the Spearman and Kendall correlation measures. *Statistical Methods & Applications*, *19*(4), 497-515.

Dakin, S. C., Tibber, M. S., Greenwood, J. A., Kingdom, F. A. A., & Morgan, M. J. (2011). A common visual metric for approximate number and density. *Proceedings of the National Academy of Sciences*, *108*(49), 19552-19557.

De Vos, T. (1992). *Test voor het vaststellen van het rekenvaardigheidsniveau der elementaire bewerkingen (automatisering) voor het basis en voortgezet onderwijs: Handleiding [Test to determine the mathematics ability level for elementary operations (automatization) in primary and secondary education: Manual].* Nijmegen, the Netherlands: Berkhout.

Dehaene, S. (1997). *The Number Sense: How the Mind creates Mathematic*. New York, NY: Oxford University Press.

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

DeWind, N. K., Adams, G. K., Platt, M. L., & Brannon, E. M. (2015). Modeling the approximate number system to quantify the contribution of visual stimulus features. *Cognition*, *142*, 247-265.

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

Ferrigno, S., Jara-Ettinger, J., Piantadosi, S. T., & Cantlon, J. F. (2017). Universal and uniquely human factors in spontaneous number perception. *Nature Communications*, *8*, 13968.

Geary, D. C. (2011). Cognitive predictors of achievement growth in mathematics: a 5-year longitudinal study. *Developmental psychology*, *47*(6), 1539.

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.

Gebuis, T., & Reynvoet, B. (2011). Generating nonsymbolic number stimuli. *Behavior Research Methods*, *43*(4), 981-986.

Gebuis, T., & Reynvoet, B. (2012). The interplay between nonsymbolic number and its continuous visual properties. *Journal of Experimental Psychology: General*, *141*(4), 642-648.

Georges, C., Hoffmann, D., & Schiltz, C. (2016). How Math Anxiety Relates to Number–Space Associations. *Frontiers in Psychology*, *07*. doi: 10.3389/fpsyg.2016.01401

Gilmore, C., Attridge, N., Clayton, S., Cragg, L., Johnson, S., Marlow, N., … Inglis, M. (2013). Individual Differences in Inhibitory Control, Not Non-Verbal Number Acuity, Correlate with Mathematics Achievement. *PLoS ONE*, *8*(6), e67374.

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

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

Guillaume, M., & Van Rinsveld, A. (2018). Comparing Numerical Comparison Tasks: A Meta-Analysis of the Variability of the Weber Fraction Relative to the Generation Algorithm. *Frontiers in Psychology*, *09*, doi: 10.3389/fpsyg.2018.01694

Krauskopf, J. (1980). Discrimination and detection of changes in luminance. *Vision Research*, *20*(8), 671-677.

Henik, A., Gliksman, Y., Kallai, A., & Leibovich, T. (2017). Size perception and the foundation of numerical processing. *Current Directions in Psychological Science*, *26*(1), 45–51.

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

Hoffmann, D., Mussolin, C., Martin, R., & Schiltz, C. (2014). The impact of mathematical proficiency on the number-space association. *PLoS ONE*, *9*(1), e85048.

Holloway, I. D., & Ansari, D. (2009). Mapping numerical magnitudes onto symbols: The numerical distance effect and individual differences in children's mathematics achievement. *Journal of Experimental Child Psychology*, *103*(1), 17-29.

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

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in Psychtoolbox-3? *Perception, 36*(14), 1–16.

Leibovich, T., Katzin, N., Harel, M., & Henik, A. (2016). From ‘sense of number’ to ‘sense of magnitude’ – The role of continuous magnitudes in numerical cognition. *Behavioral and Brain Sciences*, e164.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Martin, R., Houssemand, C., Schiltz, C., Burnod, Y., & Alexandre, F. (2008). Is there continuity between categorical and coordinate spatial relations coding? *Neuropsychologia*, *46*(2), 576-594.

Mussolin, C., Nys, J., Leybaert, J., & Content, A. (2012). Relationships between approximate number system acuity and early symbolic number abilities. *Trends in Neuroscience and Education*, *1*(1), 21-31.

Norris, J. E., Castronovo, J. (2016). Dot Display Affects Approximate Number System Acuity and Relationships with Mathematical Achievement and Inhibitory Control. *PLOS ONE*, *11(5)*, e0155543.

Norris, J. E., Clayton, S., Gilmore, C., Inglis, M., & Castronovo., J. (2018). The measurement of approximate number system acuity across the lifespan is compromised by congruency effects. *Quarterly Journal of Experimental Psychology*, doi:10.1177/1747021818779020.

Núñez, R. E. (2017). Is There Really an Evolved Capacity for Number? *Trends in Cognitive Sciences*, *21*(6), 409-424.

Nys, J., & Content, A. (2012). Judgement of discrete and continuous quantity in adults: Number counts! *The Quarterly Journal of Experimental Psychology*, *65*(4), 675-690.

Nys, J., Ventura, P., Fernandes, T., Querido, L., Leybaert, J., & Content, A. (2013). Does math education modify the approximate number system? A comparison of schooled and unschooled adults. *Trends in Neuroscience and Education*, *2*(1), 13-22.

Park, J. (2018). A neural basis for the visual sense of number and its development: A steady-state visual evoked potential study in children and adults. *Developmental Cognitive Neuroscience*, *30*, 333-343.

Pelli, D. G. (1997). The video toolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision*, *10*, 437–442.

Piazza, M., Izard, V., Pinel, P., Le Bihan, D., & Dehaene, S. (2004). Tuning Curves for Approximate Numerosity in the Human Intraparietal Sulcus. *Neuron*, *44*(3), 547-555.

Pica, P., Lemer, C., Izard, V., & Dehaene, S. (2004). Exact and Approximate Arithmetic in an Amazonian Indigene Group. *Science*, *306*(5695), 499-503.

R Core Team (2016). R: a language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. Retrieved from http://www.R-project.org

Roquet, A., & Lemaire, P. (in press). Strategy variability in numerosity comparison task: a study in young and older adults. *Open Psychology*.

Salti, M., Katzin, N., Katzin, D., Leibovich, T., & Henik, A. (2017). One tamed at a time: A new approach for controlling continuous magnitudes in numerical comparison tasks. *Behavior Research Methods*, *49*(3), 1120-1127.

Schneider, M., Beeres, K., Coban, L., Merz, S., Susan Schmidt, S., Stricker, J., & De Smedt, B. (2016). Associations of non-symbolic and symbolic numerical magnitude processing with mathematical competence: a meta-analysis. *Developmental Science*, *20(3)*, e12372.

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.

Smets, K., Moors, P., & Reynvoet, B. (2016). Effects of Presentation Type and Visual Control in Numerosity Discrimination: Implications for Number Processing? *Frontiers in Psychology*, 0*7*, doi: 10.3389/fpsyg.2016.00066

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.

Stevenson, C. E., Bergwerff, C. E., Heiser, W. J., & Resing, W. C. M. (2014). Working Memory and Dynamic Measures of Analogical Reasoning as Predictors of Children's Math and Reading Achievement: WM and Dynamic Reasoning Predict Children's School Achievement. *Infant and Child Development*, *23*(1), 51-66.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*(6), 643-662.

Townsend, J. T., & Ashby, F. G. (1978). Methods of modeling capacity in simple processing systems. In J. Castellan and F. Restle (Eds.), *Cognitive Theory Vol. III* (pp. 200-239). Hillsdale, NJ: Erlbaum Associates.

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