The interplay between non-symbolic number and its continuous visual properties revisited: Effects of mixing trials of different types

Judit Pekár[1*], Annette Kinder[1]

[1] Institute of Psychology of Learning, Department of Education and Psychology, Free University Berlin

All authors contributed equally to this work

Running title: Mixing trials in dot comparison task

*Corresponding author:

Judit Pekár, M.Sc.

Institute of Psychology of Learning

Department of Education and Psychology, Free University Berlin

Habelschwerdter Allee 45, 14195 Berlin, GERMANY

Email: judit.pekar@fu-berlin.de

Phone: +49 30 838 75953

Abstract

In the last few years, the existence of a pure number sense has been challenged. Recent studies suggest that numerosity processing is not only influenced by the number of elements in a display but also by continuous magnitudes, such as the size of the elements. The aim of our study was to replicate and extend the findings by Gebuis and Reynvoet (Gebuis & Reynvoet, 2012a), who systematically manipulated different continuous magnitudes either congruently or incongruently with discrete numerosity. We were particularly interested in finding the same pattern of congruency effects and assess its stability and robustness as this pattern indicates a complex influence of continuous magnitudes on numerosity judgements. We did so by showing stimuli of different conditions either in separate blocks or mixed together while participants solved a dot comparison task. Our results are in line with the notion that discrete number and continuous magnitudes are integrated in numerosity judgments by means of a weighing process. Moreover, our findings suggest that this integration is modified by the mode of presentation (blocked vs. mixed).

Keywords: Approximate Number System, Sensory Integration Theory, dot comparison, continuous magnitudes, numerosity

1. Introduction

The concept of numerosity is essential in everyday life. For example, we estimate and compare numerosities when we try to choose the shorter queue in the supermarket or when we try to pick the plate with more biscuits in the cafeteria. The most prominent theory in numerical cognition is the Approximate Number System (ANS). According to this theory, humans and animals share an evolutionary ancient system which is responsible for the rapid and effortless extraction of numerosities from visual and other sensory scenes (Cantlon, Platt, & Brannon, 2009; Feigenson, Dehaene, & Spelke, 2004). In particular, single-cell studies have found that number selective neurons in a dedicated fronto-parietal cortical network encode numerosity in the primate brain (for a review see Nieder, 2016). Furthermore, neuroimaging studies support the existence of the human counterpart of this network (Knops, 2017; Lyons, Ansari, & Beilock, 2015; Piazza, Izard, Pinel, Le Bihan, & Dehaene, 2004; Piazza, Pinel, Le Bihan, & Dehaene, 2007). While behavioural, neuroimaging, neurophysiological and computational studies support the existence of the ANS (Cantlon et al., 2009; Dehaene & Changeux, 1993; Feigenson et al., 2004; Gallistel & Gelman, 2000; Stoianov & Zorzi, 2012; Verguts & Fias, 2004), its exact mechanisms are less clear. In particular, it is still an open question whether and how sensory cues influence numerosity processing (Leibovich, Katzin, Harel, & Henik, 2016).

Classic theories of the ANS suggest that numerosity processing occurs independently of continuous sensory properties present in the stimuli, i.e. our estimate about the biscuits in the cafeteria is not biased by their size or by how densely they are placed on the plate. Discrete numerosity, however, is always confounded by continuous magnitudes: considering the example above, the total volume of the biscuits will be larger on the plate with more pieces, provided that all biscuits are of the same size. However, when the total biscuit volume on both plates are equal, then the size of the biscuits must be smaller on the plate with more biscuits. Even though researchers have put considerable effort into developing methods that can control the relationship between numerosity and its confounding sensory properties (Dehaene, Izard, & Piazza, 2005; Gebuis & Reynvoet, 2011; Salti, Katzin, Katzin, Leibovich, & Henik, 2016), there is always a continuous magnitude which correlates – either positively or negatively – with numerosity (see for example Smets, Moors, & Reynvoet, 2016). Thus, inspired by the ubiquitous nature of sensory properties, in the last few years, attention has been shifted to investigating what role non-numerical parameters play in numerosity processing. Based on these studies, more recent theories postulate that continuous properties influence or might even explain the processes underlying the estimation and comparison of large approximate numerosities (Clayton, Inglis, & Gilmore, 2019; Gebuis, Kadosh, & Gevers, 2016; Gevers, Kadosh, & Gebuis, 2016; Leibovich, Kallai, & Itamar, 2016; Leibovich, Katzin, et al., 2016).

Thus, on the one hand, numerous studies support pure number sense (Cantlon et al., 2009; Feigenson et al., 2004; Gallistel & Gelman, 2000; Piazza et al., 2007; Stoianov & Zorzi, 2012; Verguts & Fias, 2004) while a growing number of behavioural, neurophysiological and neuroimaging studies suggests that continuous sensory properties play an important role in numerosity processing (Gebuis & Reynvoet, 2012a, 2012b, 2013; Leibovich & Ansari, 2017; Leibovich & Henik, 2014; Salti et al., 2016; Smets, Sasanguie, Szücs, & Reynvoet, 2015; Soltész & Szűcs, 2014). Usually, in these experiments participants have to solve a dot comparison task in which two dot arrays of varying numerosities are presented to them. They are asked to indicate whether the first or the second array had contained more dots. Unbeknownst to the participants, continuous magnitudes are manipulated to create different congruency conditions. For example, on a congruent trial the array with more dots has larger visual cues, i.e. larger sized dots or a larger convex hull (larger contour around the dot array). In contrast, on an incongruent trial, the more numerous dot array has smaller visual characteristics, i.e. smaller dots or a smaller convex hull. With this method, one can manipulate several continuous magnitudes together in the same direction, e.g. dot size and convex hull are both larger on the array with more dots (congruent & congruent), or they are both smaller (incongruent & incongruent). The continuous magnitudes can also be manipulated in the opposite directions, e.g. larger dot size and a smaller convex hull on the array with more dots (congruent & incongruent). It is even possible to keep one continuous magnitude constant, while making the other congruent or incongruent. For example, the two dot arrays have the same convex hull but the more numerous one has larger dots (equated & congruent) or the less numerous one has larger dots (equated & incongruent, see also Figure 1). Comparing performance between trials with the same numerosity information but different congruency and visual cue manipulation, gives us insight into the effects of continuous magnitudes on numerosity processing. Findings of experiments which implemented this kind of sensory cue manipulation have raised the attention to the relationship between numerosity and its continuous magnitudes and questioned the existence of a pure number sense which is independent of continuous magnitudes (Gebuis & Reynvoet, 2012a, 2012b; Salti et al., 2016; Smets et al., 2015).

Sensory Integration Theory provides an alternative approach to the Approximate Number System (Gebuis et al., 2016; Gevers et al., 2016). Proponents of this theory claim that sensory cues are weighed and integrated during numerosity processing. Rather than deriving numerosity independently of sensory cues, during the so called integration procedure multiple sensory cues compete with each other in their weight given to the numerosity estimate. For example, if there is a stimulus pair which differs to a larger extent in terms of convex hull than in terms of dot diameter size, more weight might be given to the convex hull. This mechanism can explain the pattern of congruency effects found by Gebuis &

Reynvoet (2012a). Specifically, they implemented the visual cue manipulation method described above and created four different visual cue conditions. In two conditions, only a single visual cue (convex hull) or a set of visual cues (average dot diameter, aggregate surface of the dots and density, from now on referred to as diameter) were manipulated and they were either congruent or incongruent with numerosity. In the two other conditions, both visual cues were manipulated together, although in different directions (see Methods section for more detail). Consequently, neither convex hull nor any other visual cues were consistently informative since they did not correlate with numerosity throughout the task. Interestingly, Gebuis & Reynvoet (2012a) found opposite congruency effects in those two conditions, in which a single visual cue was manipulated. In the convex hull condition, performance was better when more dots occupied a larger area (other visual cues were equated). In the diameter condition the congruency effect reversed. Namely, performance was better when the array with more dots had a smaller average diameter, smaller aggregate surface and a smaller density (convex hull was equated). When all visual cues were manipulated in the same direction, the opposite effect of convex hull and the other sensory cues cancelled each other out, resulting in the absence of a congruency effect. However, when convex hull and diameter were manipulated in the opposite direction, their differential effect on the performance led to an augmented congruency effect. The authors argued that this pattern of findings support the idea that participants integrate various types of continuous dimensions, possibly by means of an additive weighing process, even when they are uninformative about numerosity.

Single cell studies suggest that neurons in the primate brain are either tuned to discrete numerosity or continuous magnitude, or they encode both discrete and continuous stimulus features. Moreover, these three types of neurons were found to be intermingled in the monkey parietal cortex (for example Tudusciuc & Nieder, 2009, for an overview see

Nieder, 2016). These findings are in line with human neuroimaging studies which showed overlapping but also segregated activations in response to continuous magnitude and numerosity in the human parietal cortex (Dormal & Pesenti, 2009; Kadosh et al., 2005; Pinel, Piazza, Le Bihan, & Dehaene, 2004). So the notion that numerosity and continuous magnitude are integrated seems to be supported by single cell and human neuroimaging studies. However, if there is an interaction between numerosity and continuous magnitudes on the basic neuronal level, then this interaction should be consistent across different continuous magnitudes. That is, some cells should respond to <u>larger</u> numerosity and <u>larger</u> convex hull while other cells should respond to <u>larger</u> numerosity as well as <u>larger</u> diameter and <u>not smaller</u> diameter. In line with this hypothesis, certain neurons in the monkey prefrontal cortex responded not only when monkeys had to select the more numerous but also when they had to select the larger stimulus. Another group of neurons responded when monkeys had to select the less numerous and the smaller stimulus. Thus, the congruency effect has been shown at a basic neural level, meaning that there are cells which are involved in both “greater” and “larger” responses and other cells that are involved in both “fewer” and “smaller” responses (Eiselt & Nieder, 2013). Taken together, these findings might provide an explanation of the interference effect between numerosity and continuous magnitudes found in behavioural investigations (Gebuis, Kenemans, de Haan, & van der Smagt, 2010; Szűcs & Soltész, 2007). However, the reversed congruency effect in the diameter condition found by Gebuis & Reynvoet (2012a), is difficult to reconcile with these findings because they suggest that displays with larger dot diameter should be judged as more numerous.

Further studies, which address the role of continuous magnitudes on numerosity processing, may add to our understanding about the interplay between discrete numerosity and continuous sensory cues. For example, it has been shown that the effect of continuous magnitudes on dot comparison performance can be altered by changes in instructions, task difficulty, stimulus duration and task context (Leibovich, Henik, & Salti, 2015; Leibovich-Raveh, Stein, Henik, & Salti, 2018). Thus, the weights that are given to continuous magnitudes are not static but seem to depend on various experimental factors. As some of these factors (such as presentation time) affect bottom-up processing whereas other factors (such as instructions) affect top-down processing, weights seem to depend on both top-down and bottom-up processes as well as on their interaction. As a result, sensory cues are processed in a highly flexible and adaptive fashion. Despite this growing body of research, it is still an open question, how exactly the weights of the continuous magnitudes are adjusted. For example, it is unclear whether the weights are adjusted within trials depending mainly on the stimulus features in the current trial or the weights are rather adapted gradually across many trials. The latter idea is supported by a study by Odic et al. who found that performance on a dot comparison task was influenced by trial order (Odic, Hock, & Halberda, 2014). Participants who were presented with easy trials first performed better on the task than participants who received difficult trials first. Thus, comparison performance can be altered through the history of previous trials.

If the weighing process indeed depends on trial history, it should make a difference whether trials of the same type are presented consecutively (blocked presentation) or whether different trials types are mixed together (mixed presentation). In the original study, Gebuis & Reynvoet (2012a) manipulated the sensory characteristics of the two dot arrays in a block-wise fashion, hence trials of the same type were shown consecutively to the participants. For example, in a certain block, the dot diameter of stimulus pairs was always manipulated in the same way (either congruently or incongruently to numerosity) whereas convex hull remained constant. In another block, it was the opposite: the convex hull of stimulus pairs was manipulated in the same way (congruent or incongruent), while dot diameter remained constant. But, assuming that the weighing process is flexible and adaptive as suggested by

Leibovich et al., and also assuming that trial history has an effect on dot comparison performance as suggested by Odic et al., it is possible that the blocked presentation mode influences the weights given to the continuous magnitudes in a particular way (Leibovich et al., 2015; Leibovich-Raveh et al., 2018; Odic et al., 2014). For example, if during the entire block the two stimuli in a trial differ with regard to convex hull whereas the other continuous cues are kept constant, the weight assigned to convex hull may increase or decrease compared to the weights of other continuous magnitudes. Hence, the specific pattern of congruency effect may at least partially be accounted for by the block-wise presentation mode. It seems even possible that the reversed congruency effect (that corresponds to a negative weight) in the diameter condition is merely a result of the blocked presentation mode and that it is not found under altered circumstances, e.g. when different trial types are mixed together. Therefore, we would like to investigate whether changing the trial history can alter the interaction between continuous magnitudes and numerosity and thus change the pattern of congruency effects reported by Gebuis & Reynvoet (2012a).

In sum, our study has two aims. First, we examine whether changing the trial history by mixing trials of different types can alter the pattern of congruency effects shown by Gebuis & Reynvoet (2012a). Second, we are particularly interested in the reversed congruency effect in the diameter condition because this effect is difficult to reconcile with single-cell studies which would suggest the opposite. Thus, we would like to investigate whether this effect remains stable when the presentation method is changed from blocked to mixed. To this end, we presented participants with different trials types in a mixed fashion and compared it with the exact replication of the original study (Gebuis & Reynvoet, 2012a).

2. Methods

2.1 Participants

Data were collected from 34 individuals (6 males, age: M=26.69, SD=5.96, range: 18 years 9 months – 36 years 5 months). All participants had normal or corrected-to-normal vision. They gave written informed consent and received course credit for their participation. Three participants (all of them females) were excluded from the data analysis because of low performance on the Dyscalculia Screener (Butterworth, 2003, see below). In total, data from 31 participants were analysed (6 males, age: M=26.26, SD=5.99, range: 18 years 9 months – 36 years 5 months).

Gebuis and Reynvoet (2012a) reported a congruency effect for convex hull with an effect size of *Cohen's* d_z=1.54 and a reversed congruency effect for diameter with an effect size of *Cohen's* d_z=.86 (calculated as *Cohen's* $d_z = t/\sqrt{n}$, according to Rosenthal, 1991). Power to find these effects with N=31 subjects (α = .05) were $1\text{-}\beta$ = 1 and .99, respectively (Faul, Erdfelder, Lang, & Buchner, 2007). A sample size of N=31 allows to detect possible moderations of congruency effects by presentation mode with an effect size of *Cohen's* d_z =.52 (i.e. medium sized-effect according to Cohen, 1988).

2.2 Stimuli and Tasks

2.2.1. Dot comparison Task. Stimuli were pairs of dot arrays which were presented consecutively to the participants. Their task was to indicate with a button press whether the first or the second image had contained more dots. Dot arrays were constructed the same way as those used by Gebuis & Reynvoet (2012a). White dots were presented on a dark grey background, dot size ranged from 0.11 degrees to 0.79 degrees in visual angle. Four visual properties were manipulated which are thought to influence numerosity judgements: (1) *convex hull* (area within the smallest contour around the dot array), (2) *aggregate surface of the dots*, (3) *average dot diameter* and (4) *density* (aggregate surface divided by convex hull).

It is important to note that three of these visual properties, namely aggregate surface, average diameter and density, are highly related. Therefore, it was not possible to differentiate between them. For example, if average dot diameter is increased, aggregate surface and density will also automatically increase while convex hull can remain constant. For this reason, aggregate surface, average dot diameter and density were manipulated together in one condition, which will be referred to as *diameter in/congruent condition* hereafter.

The manipulated visual cues could either be congruent or incongruent with numerosity. A stimulus was considered congruent if the visual property in question was greater for the array with more dots. In the *(1) convex hull in/congruent condition* the visual property manipulated was the area within the smallest contour around the dot array whereas density, aggregate surface and average diameter were kept equal across numerosities and congruency conditions. Thus, on *convex hull congruent trials* the array with more dots was associated with greater convex hull around the dot array. In contrast, on *convex hull incongruent trials*, arrays containing a larger number of dots had a smaller convex hull. Density, aggregate surface and average diameter were kept constant in this condition. That is, congruent and incongruent trials did not differ from each other with regard to these properties. In the *(2) diameter in/congruent condition* the convex hull was kept constant while the density, the average diameter of the dots and their aggregate surface were either congruent or incongruent with numerosity. In the *(3) fully in/congruent condition* both diameter and convex hull were manipulated together and they could be either congruent or incongruent with numerosity. As a result, on *fully congruent trials* all visual cues were larger in stimuli containing more dots. In contrast, on *fully incongruent trials* the more numerous dot arrays were associated with smaller visual cues. In the *(4) partially in/congruent condition* diameter and convex hull were manipulated in opposite ways: when convex hull was incongruent, diameter, density and aggregate surface were congruent and vice versa. So in short, *partially congruent trials* were congruent in diameter (as well as related visual cues) and incongruent in convex hull. Whereas, *partially incongruent trials* were incongruent in diameter (as well as related visual cues) and congruent in convex hull.

Each condition contained 192 trials half of which were congruent and the other half were incongruent (96 trials each). Every trial consisted of two dot arrays presented consecutively. One of the dot arrays always contained 24 dots, the other dot array could contain 16, 18, 20, 29, 32, 36 dots resulting in six different numerosity combinations equivalent to three different ratios. The ratios were 1:2, 1:3, 1:5 calculated as *(larger number – smaller number)/smaller number*. Trials within each condition were counterbalanced with respect to congruency, numerosity combinations and the number of dots in the first array.

It is important to note that as in the original study, the visual cues were not informative of numerosity, neither within a single cue manipulation condition nor throughout the whole task. Within each visual cue manipulation condition, half of the trials were congruent and the other half were incongruent. Moreover, the differences in continuous magnitudes and numerosity between stimulus pairs did not significantly correlate ($R<0.1$, $p>0.05$). Only in the *convex hull in/congruent condition* the aggregate surface was informative, since it was always larger for the more numerous dot array. It was not possible to control for this visual cue without making other cues informative. However, the difference in aggregate surface occurred in the same direction on congruent and incongruent trials. So any difference between trials with different congruency cannot be explained by participants utilizing aggregate surface as a cue to infer numerosity (Figure 2).

2.2.2. Dyscalculia Screener. The Dyscalculia Screener is a standardized test originally developed to assess mathematical abilities in children between the age of 6 and 14

(Butterworth, 2003). Standardized scores have been interpolated for adults and successfully used to assess mathematical proficiency in adult populations (Cappelletti et al., 2014; Cappelletti, Freeman, & Butterworth, 2011; Cappelletti & Price, 2014). The test comprises four item-timed tasks which are divided into two subscales. The capacity subscale involves a dot enumeration task and a number comparison task, the achievement subscale contains two verification tasks (mathematical addition and multiplication). Every subtask starts with detailed instructions and several practice trials with feedback. For the actual test trials no feedback is provided. Based on participants' reaction time and accuracy, individual Stanine Scores and Standard Age Scores can be computed. In previous studies, adult participants with below average results on either of the tasks of the Capacity Subscale (Dot Enumeration and/or Numerical Stroop) have been classified as dyscalculic and this result has been confirmed by further diagnostic tools (Cappelletti et al., 2014, 2011; Cappelletti & Price, 2014). Since it is still unclear whether visual cues have the same or different effect on participants with low mathematical proficiency (as raised by Leibovich, Katzin, et al., 2016), we excluded participants who had a stanine score equal to or lower than 3 on either of the two tests of the Capacity Subscale.

2.3 Procedure

First, participants completed a dot comparison task with two different presentation modes. In both presentation modes the same dot arrays were shown to the participants, the only difference was whether trials of different types were presented in a blocked or mixed fashion. During blocked presentation mode, the four visual cue conditions were presented in separate blocks and their order was counterbalanced across participants. This part of the experiment is the replication of the original study by Gebuis & Reynvoet (2012a). During mixed presentation mode, trials of all four conditions randomly alternated. The order of the presentation mode (blocked and mixed) was also counterbalanced across participants. The experiment started with detailed instructions and six practice trials with feedback. After the practice trials no feedback was provided to the participants. Each trial began with a green fixation cross presented for 500 ms on the computer screen followed by a dot array for 300 ms, a blank screen for 500 ms and the second dot array for 300 ms. Then a red fixation cross appeared on the screen and remained visible until participants pressed one of the response buttons (Figure 3). They were instructed to press the left-CTRL button if the first array was more numerous and the right-CTRL button if the second dot array was more numerous. They were asked to respond as quickly and as accurately as possible. One presentation mode consisted of four blocks each with 192 trials. In total, each participant received six practice trials and 1536 experimental trials (192 trials per block × 4 blocks × 2 presentation modes). It should be noted that participants were not aware of the different presentation modes. From their point of view, they completed eight blocks of the same dot comparison task. Hence, instructions and practice trials were presented to them only once, at the beginning of the experiment. They had the opportunity to take a break after each block. After the dot comparison task they completed the Dyscalculia Screener. The duration of the experiment was 1.5-2 hours in total.

2.4 Statistical analysis

We calculated the percentage of correct responses for each presentation mode (blocked versus mixed), type of visual cue manipulation (1: convex hull, 2: diameter, 3: fully, 4: partially in/congruent) and congruency (congruent versus incongruent). In order to determine whether we could replicate the original results and whether there was a significant difference between the blocked and mixed presentation modes, accuracy data were subjected to a 3-way repeated measures ANOVA with factors presentation mode (2), visual cue condition (4) and congruency (2). We report Greenhouse-Geisser corrected results, whenever the assumption of sphericity is violated.

3. Results

The analysis of variance revealed a main effect of presentation mode ($F(1,30)=4.69$, $p<.05$, $\eta_p^2=.14$), a main effect of visual cue condition ($F(1.81,54.29)=27.56$, $p<.001$, $\eta_p^2=.48$) and a main effect of congruency ($F(1,30)=21.25$, $p<.001$, $\eta_p^2=.42$). Main effects were qualified by a significant two-way interaction between visual cue condition and congruency ($F(1.66,49.80)=59.45$, $p<.001$, $\eta_p^2=.66$) and a three-way interaction between presentation mode, visual cue condition and congruency ($F(2.04,61.33)=6.99$, $p=.002$, $\eta_p^2=0.19$).

To follow-up the 3-way interaction, we conducted a 2-way analysis of variance with the factors visual cue conditions (4) and congruency (2), separately for the blocked and mixed presentation modes. Both ANOVAs revealed the same results: a main effect of visual cue condition (blocked: $F(2.22,66.46)=21.01$, $p<.001$, $\eta_p^2=0.41$, mixed: $F(2.10,62.94)=21.24$, $p<.001$, $\eta_p^2=0.42$), a main effect of congruency (blocked: $F(1,30)=21.85$, $p<.001$, $\eta_p^2=0.42$, mixed: $F(1,30)=19.65$, $p<.001$, $\eta_p^2=0.40$) and a significant 2-way interaction between visual cue condition and congruency (blocked: $F(2.05,61.55)=43.68$, $p<.001$, $\eta_p^2=.59$; mixed: $F(1.51,45.18)=62.13$, $p<.001$, $\eta_p^2=.67$) (Figure 4). Since the pattern of congruency effects is the same in both presentation modes (see Figure 4), the three-way interaction of presentation mode × visual cue conditions × congruency dominantly reflects a moderation in the congruency effects. Therefore, we report four two-way ANOVAs ([blocked vs. mixed] x [congruent vs. incongruent]) for every visual cue condition. We also report follow up t-tests for the single congruency effects (congruent vs. incongruent), separately for each presentation mode and visual cue condition.

3.1 Convex hull in/congruent condition

The 2-way analysis of variance conducted for the convex *hull in/congruent visual cue condition* revealed a significant main effect of congruency (F(1, 30)=88.52, p<.001, η_p^2=.75) and a significant interaction of presentation mode and congruency (F(1,30)=19.55, p<0.001, η_p^2=.40). The main effect of congruency was due to better performance on congruent trials (M=87.2, SD=1.26) than on incongruent trials (M=74.65, SD=1.72, see also Figure 4). Participants made significantly more correct responses when the array with more dots had a larger convex hull than when it had a smaller convex hull (all other visual cues equated). The interaction of presentation mode and congruency was due to a significantly larger congruency effect in the mixed presentation mode (congruent: M=88.71, SD=6.93, incongruent: M=72.45, SD=11.31) than in the blocked presentation mode (congruent: M=85.69, SD=8.29, incongruent: M=76.85, SD=9.71). The congruency effects ([congruent vs. incongruent]) were significant in both presentation modes (blocked: t(30)=5.87, p<.001, Cohen's d= 1.06, mixed: t(30)=9.89, p<.001, Cohen's d=1.78)

3.2 Diameter in/congruent condition

The ANOVA showed a significant main effect of presentation mode (F(1,30)=6.72, p<.05, η_p^2=.18), a significant main effect of congruency (F(1,30)=21.95, p<.001, η_p^2=.42) and a significant interaction of presentation mode and congruency (F(1,30)=7.52, p<.05, η_p^2=.20, see also Figure 4). The main effect of presentation mode was the result of better performance in the blocked (M=77.52, SD=1.68) than in the mixed presentation mode (M= 74.90, SD=1.73). The main effect of congruency was due to lower performance on congruent (M=64.80, SD=3.71) than on incongruent trials (M=87.62, SD=1.85). These results show that participants gave more correct responses when the array with more dot was associated with smaller visual cue (incongruent trials) which is in line with the results of the original study by Gebuis & Reynvoet (2012a). The interaction was again the result of a smaller congruency effect in the blocked (congruent: *M*=67.55, *SD*=3.67, incongruent: *M*=87.60 *SD*=1.90) than in the mixed presentation mode (congruent: *M*=62.16 *SD*=3.93, incongruent: *M*=87.63, *SD*=1.96). Follow-up t-tests showed that the congruency effects ([congruent vs. incongruent]) were significant during both presentation modes (blocked: *t*(30)=-4.22, *p*<.001, Cohen's *d*=-.76, mixed: *t*(30)=-4.95, *p*<.001, Cohen's *d*=-.89).

3.3 Fully in/congruent condition

The ANOVA did not yield any significant results (presentation mode: *F*(1,30)=2.96, *p*=.10, η_p^2=.09, congruency: F(1,30)=1.83, p=.19, η_p^2=.06, presentation mode × congruency: *F*(1,30)=1.78, *p*=.19, η_p^2=.06). T-tests on the congruency effects have not revealed any significant results either (blocked: *t*(30)=-.87, *p*=.39, Cohen's *d*=-.16, mixed: *t*(30)=-1.6, *p*=.12, Cohen's *d*=-.29).

3.4 Partially in/congruent condition

The analysis conducted for the *partially in/congruent condition* revealed a significant main effect of congruency (*F*(1,30)=79.66, *p*<.001, η_p^2=.73) and a significant interaction of presentation mode and congruency (*F*(1,30)=4.75, *p*<.05, η_p^2=.14). The main effect of congruency was a result of lower performance on *partially congruent trials* (diameter congruent, convex hull incongruent), (*M*=49.56, *SD*=4.15) than on *partially incongruent trials* (diameter incongruent, convex hull congruent), (*M*=92.00, *SD*=1.16). Again, the interaction of presentation mode and congruency was due to an increase in congruency effect from blocked (congruent: *M*=50.47, *SD*=4.14; incongruent: *M*=91.03, *SD*=1.42) to mixed presentation mode (congruent: *M*=48.66, *SD*=4.29; incongruent: *M*=92.77, *SD*=1.06). The congruency effects were significant in both presentation modes (blocked: *t*(30)=-8.48, *p*<.001, Cohen's *d*=-1.52, mixed: *t*(30)=-4.99, *p*<.001, Cohen's *d*=-1.63) Altogether, these findings indicate that participants gave more correct responses, when the array with more dots had a larger convex hull but smaller average diameter, aggregate surface and density *(partially incongruent trials)*. This pattern of results corresponds to the original findings by Gebuis & Reynvoet (2012a).

4. Discussion

The current study had two objectives. First, we examined whether changing the trial history by mixing trials of different types can alter the pattern of congruency effects shown by Gebuis & Reynvoet (2012a). Second, we were particularly interested in the reversed congruency effect in the diameter condition, when larger dots were estimated less numerous because this result is difficult to reconcile with findings from single-cell studies which would suggest the opposite (Nieder, 2016). We wanted to determine if the reversed diameter effect in the original study was merely a by-product of the block-wise presentation mode or whether it could be replicated when trials of different types are mixed together. For this purpose, we showed participants different trial types in a mixed fashion and compared it to the exact replication of the original study, where trials with different types of visual cue manipulation methods were presented in separate blocks.

We replicated the pattern of congruency effects found by Gebuis & Reynvoet (2012a). As in the original study, the same numerosities were used in all visual cue conditions, so any difference between congruent and incongruent trials can only be attributed to how the visual cues were manipulated within that condition. In the *convex hull in/congruent condition*, performance was better on congruent trials, i.e. when the more numerous dot array occupied a larger area (other visual cues were equated). We were also able to replicate the reversed congruency effect in the *diameter in/congruent condition*. Performance was better on incongruent trials, when the array with more dots had smaller average diameter, smaller aggregate surface and smaller density (convex hull was equated).

When all visual cues were manipulated in the same direction in the *fully in/congruent condition*, no congruency effect was found. In the *partially in/congruent condition*, however, when the visual cues were manipulated in the opposite direction, the congruency effect increased. Gebuis & Reynvoet (2012a) draw the conclusion that when all visual cues are manipulated in the same direction *(fully in/congruent condition)*, the opposite effects of convex hull and diameter results in an attenuated congruency effect. However, when convex hull and dot diameter are manipulated in different directions *(partially in/congruent condition)*, an increase in the congruency effect is induced. Taken together, we were able to completely replicate these findings of Gebuis & Reynvoet (2012a), including the reversed congruency effect in the *diameter in/congruent condition*. In contrast to our study, other studies using the same method to generate non-symbolic stimuli, included only dot arrays of the *fully* and the *partially in/congruent conditions* (Gilmore et al., 2013; Gilmore, Cragg, Hogan, & Inglis, 2016; Leibovich, Vogel, Henik, & Ansari, 2016; Smets et al., 2015; Szucs, Nobes, Devine, Gabriel, & Gebuis, 2013). Hence, our study is the first to reproduce the congruency effect in the *convex hull* and the reversed congruency effect in the *diameter in/congruent condition*.

We also compared performance and congruency effects between blocked and mixed presentation modes. As mentioned above, the pattern of congruency effects did not differ between them. However, congruency effects were greater when trials of different types were mixed together and this increase in congruency effect was significant in the *convex hull, diameter* and *partially in/congruent* conditions. It is important to note that these differences between the two presentation modes can only be attributed to how trials were presented to the participants (block-wise vs. mixed together), since the same dot arrays were shown during both tasks. The fact that the same pattern of congruency effects was found when trials of different types were mixed together supports the notion that convex hull and diameter have opposite effects on numerosity processing. It also shows that this is a stable effect and not simply a byproduct of the block-wise stimulus presentation method. Moreover, the increased congruency effects demonstrate that reliance on visual cues became larger when trials of different types were mixed together. Taken together, these results support that continuous magnitudes play an important role in numerosity processing and indicate that their role is not as straightforward as previously thought.

Specifically, it is noteworthy that whereas convex hull and numerosity are combined in a "greater is more" fashion, diameter and numerosity are combined in a "smaller is more" fashion. If the integration of numerosity and continuous magnitudes occurred on a basic neural level, one would expect a "greater is more" rule with all continuous magnitudes. It is possible that the reversed congruency effect in the diameter condition is a result of past experience and reflects a learning process about the relationship between numerical and non-numerical parameters. In real life, there tends to be a negative correlation between item diameter and numerosity when other sensory cues are equal. For example, when one of two equal-sized baskets is filled with apples and the other one with peas, then there must be in total more pieces of peas than apples. It is possible that participants have learned this negative relationship between the size of the items and their numerosity and applied what they have learned in the experiment when comparing dot patterns. This hypothesis is supported by the stability of the reversed diameter effect: it is not only present across different trial presentation modes (blocked vs. mixed) but also across different dot generation protocols. In fact, the effect has been replicated by studies using the Panamath method for generating dot comparison stimuli (Clayton, Gilmore, & Inglis, 2015; Norris, Clayton, Gilmore, Inglis, & Castronovo, 2019). This is an important finding as some concerns have been raised about the replicability of findings in dot comparison tasks across different dot generation protocols (Inglis & Gilmore, 2014; Smets et al., 2015). There is only a single study in which larger dot diameter *increased* (rather than decreased) perceived numerosity in dot patterns (Salti et al., 2016). Yet, this contradiction can be resolved when one looks at the numerosity ranges used in this study. Whereas our study as well as all other studies finding a reversed diameter effect presented numerosities from the estimation range, Salti et al. (2016) used numerosities from the subitizing range (2-4). Thus, it seems that within subitizing range dot diameter is integrated during numerosity processing the same way as other continuous magnitudes and this pattern is reversed in the estimation range. This differential effect of diameter on numerosity judgements further supports the notion that numerosities in the estimation and subitizing ranges are processed differently (Feigenson et al., 2004; Hyde & Spelke, 2009, 2011, 2012; Plodowski, Swainson, Jackson, Rorden, & Jackson, 2003). A possible explanation is that diameter is combined with numerosity in a “greater is more” fashion in the subitizing range which is in line with findings of single-cell studies while the reversed congruency effect (“smaller is more”) in the estimation range is possibly a result of past experience and reflects a learning process about the relationship between numerical and non-numerical parameters. This possibility, that learning is crucial in the integration of visual cues and numerosity, has been raised by several researchers (Gebuis et al., 2016; Leibovich, Katzin, et al., 2016; Mix, Huttenlocher, & Levine, 2002).

Whereas the overall pattern of congruency effects was identical in both presentation modes, we could show that the influence of continuous magnitudes on performance increased when different trial types were presented in a mixed fashion. When a trial was preceded by trials of different types (as during mixed presentation mode) participants relied more on sensory cues than they did when the trial was preceded by trials of the same type (as during blocked presentation mode). In terms of Sensory Integration Theory this finding can be interpreted as an increase in weights given to the continuous magnitudes. Thus, our results add to the growing body of evidence suggesting that these weights are dynamic and applied in an adaptive and flexible way (Leibovich et al., 2015; Leibovich, Kallai, et al., 2016; Leibovich, Katzin, et al., 2016; Leibovich-Raveh et al., 2018). As mentioned before, different instructions, task difficulty, exposure duration and task context can alter the effect of continuous magnitudes on dot comparison performance. The differential effect of continuous magnitudes between blocked and mixed presentation modes in our study demonstrates that the weighing process depends not only on these factors but also on experiences participants have made with non-symbolic numerosities, including very recent ones in previous experimental trials.

Sensory Integration Theory assumes that the process of weighing the different continuous magnitudes when estimating the number of dots in a pattern is essentially a perceptual one. It is also possible, however, that changing the trial history has an impact on numerosity processing at some point later in the processing line. Leibovich et al. (Leibovich, Kallai, et al., 2016; Leibovich, Katzin, et al., 2016) emphasize the role of cognitive control abilities in numerosity processing: integration is necessary to allow us to use the natural correlation between continuous magnitudes and numerosity (e.g. congruent trials) but inhibition is required to suppress our bias to process the visual cues when the natural correlations are violated (e.g. incongruent trials). Thus, any detrimental influence of the visual cues on numerosity judgements may be a result of a deficient inhibition process. In line with this idea, studies found that both inhibition and integration play a role in numerosity processing (Cappelletti, Pikkat, Upstill, Speekenbrink, & Walsh, 2015; Clayton & Gilmore, 2015; Fuhs & McNeil, 2013; Gilmore et al., 2013). An inhibition account of our results would suggest that it is more difficult to inhibit continuous magnitudes when different trial types are mixed together. The inhibition account may also provide an explanation for the reversed diameter effect: diameter is encoded the same way on the neuronal level as other continuous magnitudes but inhibition is required to overcome biases induced by dot size when making numerosity comparisons. The exact role of inhibition should be further investigated possibly by testing individuals of different age groups that have different levels of inhibitory control.

It should also be pointed out that the increased congruency effects during the mixed compared to the blocked presentation mode have practical implications. Dot comparison tasks are the most dominant methods to assess the acuity of the Approximate Number System. Although methods have been developed which aim to quantify and exclude the influence of visual cues on the ANS acuity measurement (DeWind, Adams, Platt, & Brannon, 2015), our study shows that not only the visual cues but also the presentation method may increase the reliance on sensory properties. This in turn might lead to an incorrect estimation of the ANS acuity. Our data show that this might especially be problematic when trials of different types are mixed together (e.g. Fazio, Bailey, Thompson, & Siegler, 2014; Gomez et al., 2015; Tokita & Ishiguchi, 2013). Consequently, the intention of reducing the reliance of continuous magnitude by mixing trials of different visual cue manipulation might have, in turn, led to the exact opposite – namely that the reliance on them has increased.

The role of continuous magnitudes in numerosity processing is a subject of debate in the current literature. Even though, very recent neurophysiological studies emphasize the superiority of numerosity over continuous magnitudes very early in the processing stream (Park, DeWind, Woldorff, & Brannon, 2015), studies investigating the role of continuous magnitudes – including the present one – confirm repeatedly, that these indeed have a great impact on numerosity judgements. How great and complex this impact might be, can be best illustrated by performance in the partial condition. When visual cues follow a specific pattern, performance is close to a ceiling effect (i.e. *partially incongruent trials* = convex hull congruent, diameter incongruent) but when visual cues are combined in the opposite manner, surprisingly performance drops close to chance level (i.e. *partially congruent trials* = convex hull incongruent, diameter congruent). This pattern shows that visual cues have a massive effect on the perception of numerosities if they are manipulated in a certain manner. The fact that these same patterns of congruency effects were found during blocked and mixed presentation modes indicates that visual cues have a large and stable impact on numerosity judgments at some point in the processing line.

In sum, our results further support the notion that people integrate various continuous magnitudes when making numerosity judgements. Moreover, this integration process seems to be complex and adaptive. It seems to depend not only on the type of continuous magnitudes but also on experiences participants have made with non-symbolic numerosities, including very recent ones in previous experimental trials. It still needs to be investigated whether this process is essentially a perceptual one, that includes weighing of sensory cues, or whether higher-order processes, such as inhibition, are also involved.

References

Butterworth, B. (2003). *Dyscalculia screener*. Citeseer.

Cantlon, J. F., Platt, M. L., & Brannon, E. M. (2009). Beyond the number domain. *Trends in Cognitive Sciences*, *13*(2), 83–91.

Cappelletti, M., Chamberlain, R., Freeman, E. D., Kanai, R., Butterworth, B., Price, C. J., & Rees, G. (2014). Commonalities for numerical and continuous quantity skills at temporo-parietal junction. *Journal of Cognitive Neuroscience*, *26*(5), 986–999.

Cappelletti, M., Freeman, E. D., & Butterworth, B. L. (2011). Time processing in dyscalculia. *Frontiers in Psychology*, *2*(364), 2.

Cappelletti, M., Pikkat, H., Upstill, E., Speekenbrink, M., & Walsh, V. (2015). Learning to integrate versus inhibiting information is modulated by age. *Journal of Neuroscience*, *35*(5), 2213–2225.

Cappelletti, M., & Price, C. J. (2014). Residual number processing in dyscalculia. *NeuroImage: Clinical*, *4*, 18–28.

Clayton, S., & Gilmore, C. (2015). Inhibition in dot comparison tasks. *Zdm*, *47*(5), 759–770.

Clayton, S., Gilmore, C., & Inglis, M. (2015). Dot comparison stimuli are not all alike: The effect of different visual controls on ANS measurement. *Acta Psychologica*, *161*, 177–184.

Clayton, S., Inglis, M., & Gilmore, C. (2019). Developmental differences in approaches to nonsymbolic comparison tasks. *Quarterly Journal of Experimental Psychology*, *72*(3), 436–445.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences. Abingdon*. United Kingdom: Routledge.

Dehaene, S., & Changeux, J.-P. (1993). Development of elementary numerical abilities: A neuronal model. *Journal of Cognitive Neuroscience*, *5*(4), 390–407.

Dehaene, S., Izard, V., & Piazza, M. (2005). *Control over non-numerical parameters in numerosity experiments*.

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.

Dormal, V., & Pesenti, M. (2009). Common and specific contributions of the intraparietal sulci to numerosity and length processing. *Human Brain Mapping*, *30*(8), 2466–2476.

Eiselt, A.-K., & Nieder, A. (2013). Representation of abstract quantitative rules applied to spatial and numerical magnitudes in primate prefrontal cortex. *Journal of Neuroscience*, *33*(17), 7526–7534.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191.

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

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

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

Gallistel, C. R., & Gelman, R. (2000). Non-verbal numerical cognition: From reals to integers. *Trends in Cognitive Sciences*, *4*(2), 59–65.

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

Gebuis, T., Kenemans, J. L., de Haan, E. H., & van der Smagt, M. J. (2010). Conflict processing of symbolic and non-symbolic numerosity. *Neuropsychologia*, *48*(2), 394–401.

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

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

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

Gebuis, T., & Reynvoet, B. (2013). The neural mechanisms underlying passive and active processing of numerosity. *NeuroImage*, *70*, 301–307.

Gevers, W., Kadosh, R. C., & Gebuis, T. (2016). Sensory integration theory: An alternative to the approximate number system. In *Continuous issues in numerical cognition* (pp. 405–418). Elsevier.

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.

Gomez, A., Piazza, M., Jobert, A., Dehaene-Lambertz, G., Dehaene, S., & Huron, C. (2015). Mathematical difficulties in developmental coordination disorder: Symbolic and nonsymbolic number processing. *Research in Developmental Disabilities*, *43*, 167–178.

Hyde, D. C., & Spelke, E. S. (2009). All numbers are not equal: An electrophysiological investigation of small and large number representations. *Journal of Cognitive Neuroscience*, *21*(6), 1039–1053.

Hyde, D. C., & Spelke, E. S. (2011). Neural signatures of number processing in human infants: Evidence for two core systems underlying numerical cognition. *Developmental Science*, *14*(2), 360–371.

Hyde, D. C., & Spelke, E. S. (2012). Spatiotemporal dynamics of processing nonsymbolic number: An event-related potential source localization study. *Human Brain Mapping*, *33*(9), 2189–2203. https://doi.org/10.1002/hbm.21352

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

Kadosh, R. C., Henik, A., Rubinsten, O., Mohr, H., Dori, H., van de Ven, V., … Linden, D. E. (2005). Are numbers special?: The comparison systems of the human brain investigated by fMRI. *Neuropsychologia*, *43*(9), 1238–1248.

Knops, A. (2017). Probing the neural correlates of number processing. *The Neuroscientist*, *23*(3), 264–274.

Leibovich, T., & Ansari, D. (2017). Accumulation of non-numerical evidence during nonsymbolic number processing in the brain: An fMRI study. *Human Brain Mapping*, *38*(10), 4908–4921.

Leibovich, T., & Henik, A. (2014). Comparing performance in discrete and continuous comparison tasks. *The Quarterly Journal of Experimental Psychology*, *67*(5), 899–917.

Leibovich, T., Henik, A., & Salti, M. (2015). Numerosity processing is context driven even in the subitizing range: An fMRI study. *Neuropsychologia*, *77*, 137–147.

Leibovich, T., Kallai, A. Y., & Itamar, S. (2016). What do we measure when we measure magnitudes? In *Continuous issues in numerical cognition* (pp. 355–373). Elsevier.

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*, 1–62.

Leibovich, T., Vogel, S. E., Henik, A., & Ansari, D. (2016). Asymmetric processing of numerical and nonnumerical magnitudes in the brain: An fMRI study. *Journal of Cognitive Neuroscience*, *28*(1), 166–176.

Leibovich-Raveh, T., Stein, I., Henik, A., & Salti, M. (2018). Number and Continuous Magnitude Processing Depends on Task Goals and Numerosity Ratio. *Journal of Cognition*, *1*(1).

Lyons, I. M., Ansari, D., & Beilock, S. L. (2015). Qualitatively different coding of symbolic and nonsymbolic numbers in the human brain. *Human Brain Mapping*, *36*(2), 475–488.

Mix, K. S., Huttenlocher, J., & Levine, S. C. (2002). Multiple cues for quantification in infancy: Is number one of them? *Psychological Bulletin*, *128*(2), 278.

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

Norris, J. E., Clayton, S., Gilmore, C., Inglis, M., & Castronovo, J. (2019). The measurement of approximate number system acuity across the lifespan is compromised by congruency effects. *Quarterly Journal of Experimental Psychology*, *72*(5), 1037–1046.

Odic, D., Hock, H., & Halberda, J. (2014). Hysteresis affects approximate number discrimination in young children. *Journal of Experimental Psychology: General*, *143*(1), 255.

Park, J., DeWind, N. K., Woldorff, M. G., & Brannon, E. M. (2015). Rapid and direct encoding of numerosity in the visual stream. *Cerebral Cortex*, *26*(2), 748–763.

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

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

Pinel, P., Piazza, M., Le Bihan, D., & Dehaene, S. (2004). Distributed and overlapping cerebral representations of number, size, and luminance during comparative judgments. *Neuron*, *41*(6), 983–993.

Plodowski, A., Swainson, R., Jackson, G. M., Rorden, C., & Jackson, S. R. (2003). Mental representation of number in different numerical forms. *Current Biology*, *13*(23), 2045–2050.

Rosenthal, R. (1991). *Meta-Analytic Procedures for Social Research* (Revised, Subsequent). Newbury Park: SAGE Publications Inc.

Salti, M., Katzin, N., Katzin, D., Leibovich, T., & Henik, A. (2016). One tamed at a time: A new approach for controlling continuous magnitudes in numerical comparison tasks. *Behavior Research Methods*, 1–8.

Smets, K., Moors, P., & Reynvoet, B. (2016). Effects of presentation type and visual control in numerosity discrimination: Implications for number processing? *Frontiers in Psychology*, *7*, 66.

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.

Soltész, F., & Szűcs, D. (2014). Neural adaptation to non-symbolic number and visual shape: An electrophysiological study. *Biological Psychology*, *103*, 203–211.

Stoianov, I., & Zorzi, M. (2012). Emergence of a'visual number sense'in hierarchical generative models. *Nature Neuroscience*, *15*(2), 194.

Szucs, 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.

Szűcs, D., & Soltész, F. (2007). Event-related potentials dissociate facilitation and interference effects in the numerical Stroop paradigm. *Neuropsychologia*, *45*(14), 3190–3202.

Tokita, M., & Ishiguchi, A. (2013). Effects of perceptual variables on numerosity comparison in 5–6-year-olds and adults. *Frontiers in Psychology*, *4*, 431.

Tudusciuc, O., & Nieder, A. (2009). Contributions of primate prefrontal and posterior parietal cortices to length and numerosity representation. *Journal of Neurophysiology*, *101*(6), 2984–2994.

Verguts, T., & Fias, W. (2004). Representation of number in animals and humans: A neural model. *Journal of Cognitive Neuroscience*, *16*(9), 1493–1504.

Figures and figure legends

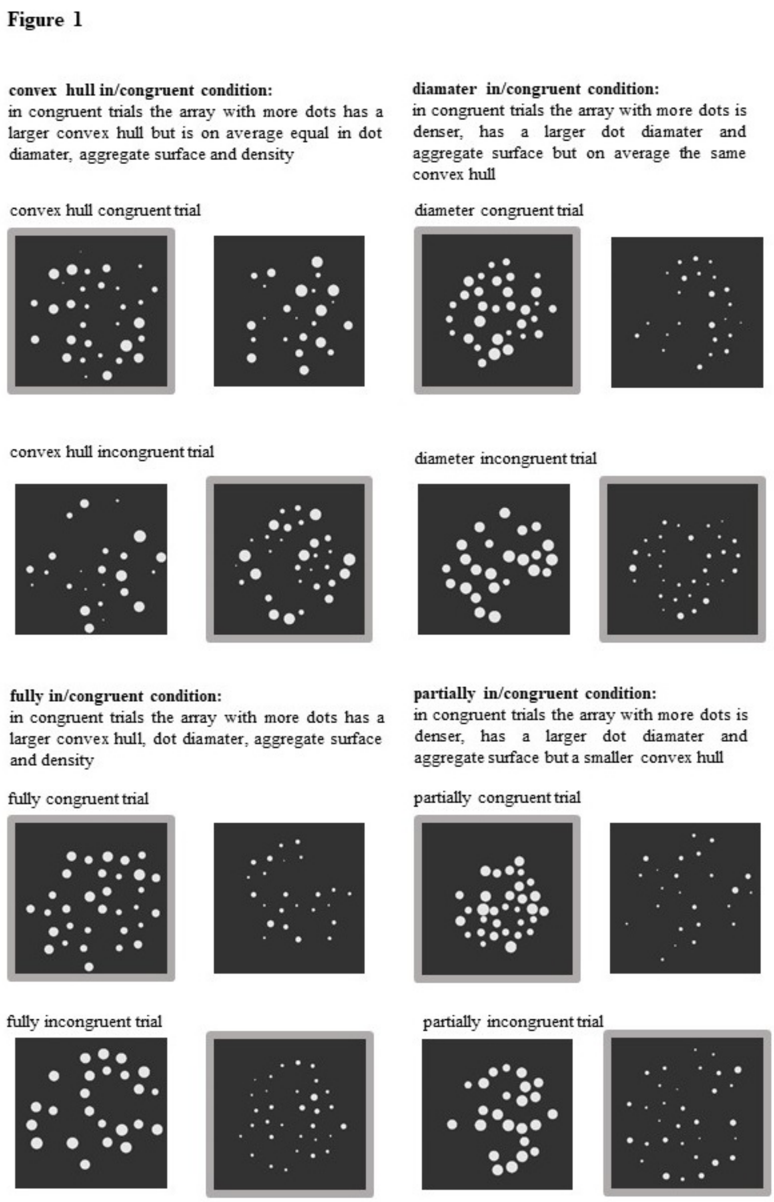

Figure 1: Examples of congruent and incongruent trials. For each of the four conditions one congruent and one incongruent stimulus pair is shown. The more numerous stimulus is marked with grey border.

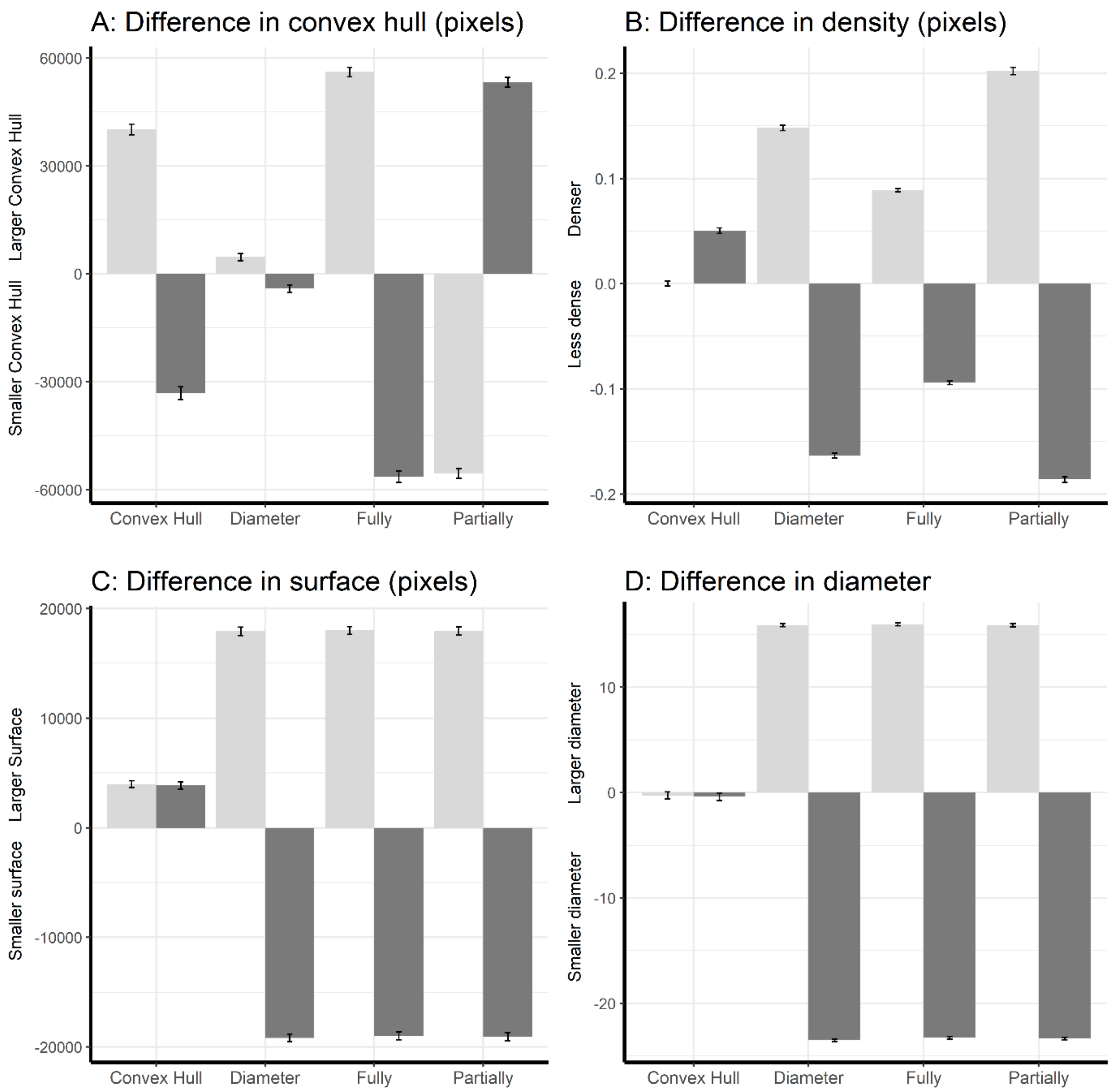

Figure 2: The difference in visual properties of all trials for each visual cue condition separately. Each panel depicts the difference in a certain visual property for the stimuli containing more dots relative to the stimuli containing less dots. They grey and the black bars represent the difference in visual properties of each number pair for the congruent and incongruent trials respectively, calculated as the visual cue of the larger number minus the visual cue of the smaller number. Error bars are displaying standard error of the mean.

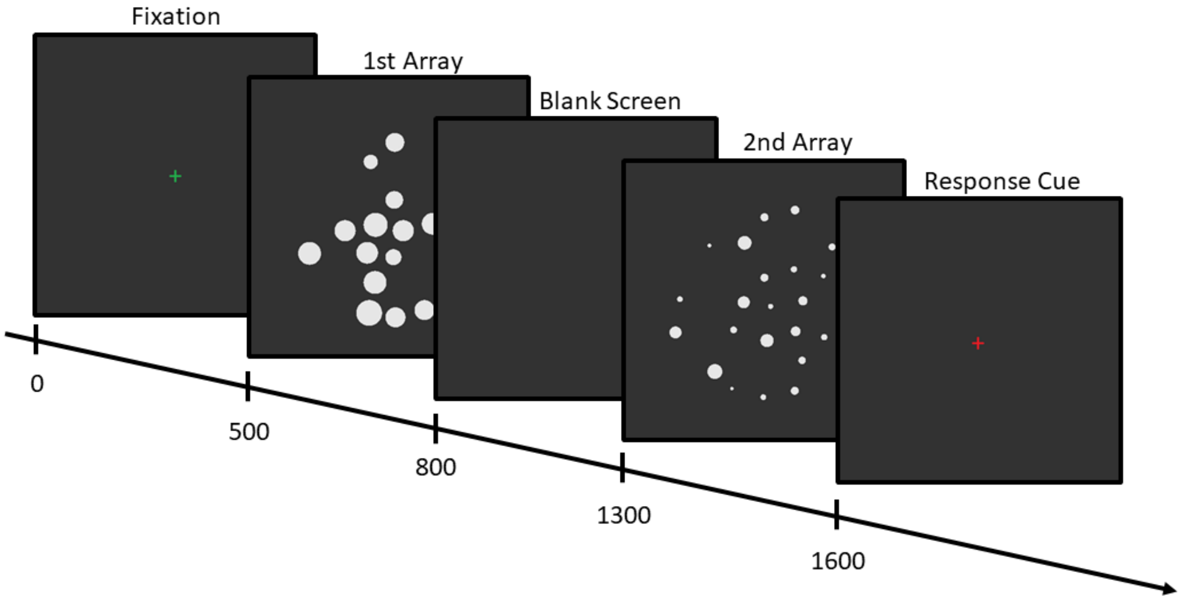

Figure 3: Experimental design with timing information.

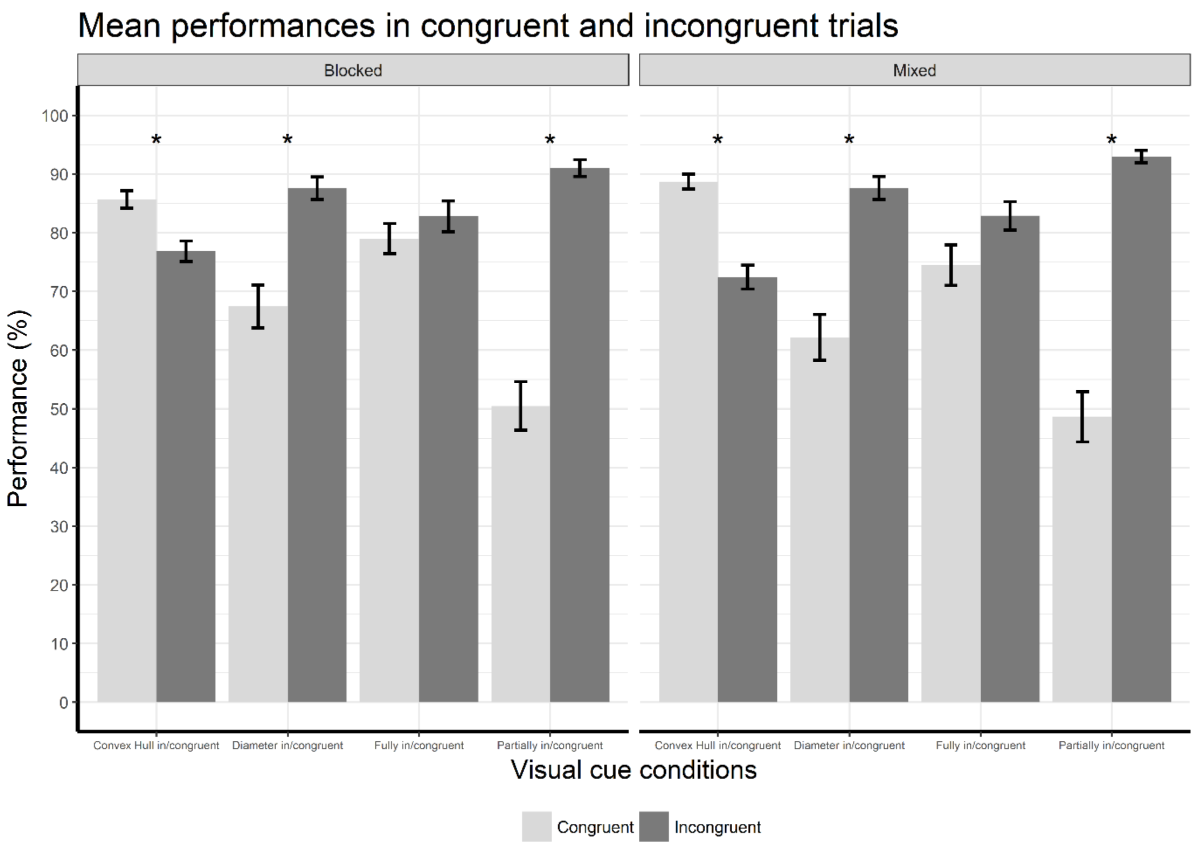

Figure 4: Performance (%) for each congruency and visual cue condition separately for the blocked and mixed presentation modes. In both presentation modes congruency effects were found in all conditions except the fully in/congruent condition. Light and dark grey bars represent congruent and incongruent trials, respectively. They also correspond to the light and dark grey bars in Figure 2. The asterisks represent significant results ($p_{adj} < .05$) Please note that *partially congruent trials* were congruent in diameter (as well as related visual cues) and incongruent in convex hull. Whereas, *partially incongruent trials* were incongruent in diameter (as well as related visual cues) and congruent in convex hull. Error bars are displaying standard error of the mean.