

Developmental differences in approaches to nonsymbolic comparison tasks

Word count: 5610

Sarah Clayton

Department of Health Sciences, University of Leicester, UK

Matthew Inglis and Camilla Gilmore

Mathematics Education Centre, Loughborough University, UK

Author Note

This research was supported by a Royal Society Dorothy Hodgkin Fellowship (C.G.), and a Royal Society Worshipful Company of Actuaries Research Fellowship (M.I.).

Correspondence concerning this paper should be addressed to Sarah Clayton, Department of Health Sciences, University of Leicester, Leicester, LE1 7HA, United Kingdom. Email: sec56@le.ac.uk. Telephone: +44 (0)116 252 3210.

Abstract

Nonsymbolic comparison tasks are widely used to measure children's and adults' Approximate Number System (ANS) acuity. Recent evidence has demonstrated that task performance can be influenced by changes to the visual characteristics of the stimuli, leading some researchers to suggest it is unlikely that an ANS exists that can extract number information independently of the visual characteristics of the arrays. Here we analysed 124 children's and 120 adults' dot comparison accuracy scores from three separate studies to investigate individual and developmental differences in how numerical and visual information contribute to nonsymbolic numerosity judgements. We found that, in contrast to adults, the majority of children did not use numerical information over and above visual cue information to compare quantities. This finding was consistent across different studies. The results have implications for research on the relationship between dot comparison performance and formal mathematics achievement. Specifically, if most children's performance on dot comparison tasks can be accounted for without the involvement of numerical information, it seems unlikely that observed correlations with mathematics achievement stem from ANS acuity alone.

Keywords: Approximate Number System, nonsymbolic comparison task, visual cues, numerical cognition

Developmental differences in approaches to nonsymbolic comparison tasks

Introduction

Throughout the last decade, nonsymbolic dot comparison tasks have been used as a tool to measure numerical representations of quantity. Such tasks are designed to assess an individual's Approximate Number System (ANS) acuity through multiple comparisons of brief displays of dot arrays. Dot comparison tasks have been assumed to measure ANS processing because performance follows the Weber-Fechner law: participants' accuracy declines as the ratio between the to-be-compared arrays approaches 1 (Dehaene, 1997). The dominant 'number sense' model of the ANS (Barth et al., 2005; Dehaene, 1997; Feigenson, Dehaene, & Spelke, 2004) proposes that imprecise and abstract representations of numerical magnitude are generated when comparing nonsymbolic arrays, and that these representations are formed independently of non-numerical continuous magnitudes (Feigenson et al., 2004; Leibovich et al., 2016). The number sense account postulates that only the ratio of the to-be-compared numerosities and the individual's ANS acuity should influence dot comparison judgements.

Recently, there has been an increase in research attention paid to the reliability and validity of dot comparison tasks. This interest, at least in part, stems from the high-profile debate around the relationship between ANS acuity and formal mathematics ability. Many studies have found a significant correlation between nonsymbolic dot comparison performance and symbolic mathematics achievement across the lifespan (e.g. Halberda, Ly, Wilmer, Naiman, & Germine, 2012; see Chen & Li, 2014, and Schneider et al. 2016, for meta-analyses). Nevertheless, there are also a number of studies that have not found a significant correlation between dot comparison performance and mathematics achievement, in both adults (e.g. Inglis, Attridge, Batchelor & Gilmore, 2011) and children (e.g. Holloway & Ansari, 2009; see De Smedt, Noël, Gilmore, & Ansari, 2013, for a review).

A contributing factor to the inconsistencies in the literature may be the validity of the tasks. Recent evidence has shown that dot comparison tasks are not pure measures of ANS acuity, and that the visual characteristics of the dot array stimuli also substantially influence

judgements (Clayton & Gilmore, 2015; Clayton, Inglis, & Gilmore, 2015; Fuhs & McNeil, 2013; Gebuis & Reynvoet, 2012; Gilmore et al., 2013; Smets, Sasanguie, Szűcs, & Reynvoet, 2015; Szűcs, Nobes, Devine, Gabriel, & Gebuis, 2013). To ensure that the visual characteristics of dot array stimuli are not informative of the number of dots, researchers attempt to control potential cues such as the average dot size, the cumulative surface area of the dots, the density of the dots, and the convex hull of the arrays (the smallest contour surrounding the dots). These manipulations are intended to ensure that it is not possible for an individual to perform significantly above chance across a set of trials by focusing on a particular visual cue, without the involvement of numerical processing. There are several approaches to constructing dot array stimuli and some methods employ more rigorous controls than others by manipulating multiple features of the arrays simultaneously (Clayton et al., 2015; Smets et al., 2015). Gebuis and Reynvoet (2011) criticised methods that only control for a single visual cue at a time and found that participants are likely to rely on multiple visual cues and switch between them depending on the trial characteristics. Gebuis and Reynvoet provided an example of a trial where one visual cue, e.g. average dot size, is equated across the two arrays and therefore uninformative of number. In such a case, they suggested that participants are likely to switch their focus to an uncontrolled visual cue, e.g. cumulative surface area, which covaries with number. From this, Gebuis and Reynvoet (2012) suggested that the existence of an ANS that is independent of visual cues is unlikely.

Indeed, researchers have since developed alternative theories of nonsymbolic numerosity processing. Leibovich et al. (2016) argued that the correlation between numerosity and continuous magnitudes, which is prominent in daily life, is a serious barrier to researchers who aim to assess non-symbolic numerosity processing in isolation from visual cues. Leibovich and colleagues proposed that both numerosities and continuous magnitudes are processed holistically when judging nonsymbolic quantities. According to this account, participants completing a dot comparison task will integrate all available numerical and non-numerical cues in order to make their decision, rather than focussing on a single magnitude. In line with this, Gebuis and Reynvoet (2012) pointed out that, given the natural correlation

between number and visual cues, a system which is unique to numerical processing would be inefficient.

Others have similarly recognised the influence of non-numerical stimulus features on numerical judgements, particularly when visual cues are misleading, resulting in the development of the competing processes account (Clayton & Gilmore, 2015; Fuhs & McNeil, 2013; Gilmore et al., 2013; Nys & Content, 2012). According to this account, inhibitory control skills play an important role in nonsymbolic comparison tasks in cases where the size of the visual characteristics (e.g. dot size or convex hull) is incongruent with the numerosity represented (i.e. the side with larger visual cues represents the numerically smaller set). In these cases, unhelpful visual cues may interfere with numerosity processing, thus requiring inhibition skills to ignore the irrelevant and misleading non-numerical information and focus on making a judgement solely based on numerosity. Support for this account comes from several dot comparison studies which have demonstrated that participants perform significantly more accurately on trials where visual cues are congruent with numerosity in comparison to incongruent trials (Barth et al., 2006; Clayton & Gilmore, 2015; Clayton et al., 2015; Gebuis, Kadosh, de Haan, & Henik, 2009; Gilmore et al., 2013; Hurewitz et al., 2006; Nys & Content, 2012), akin to performance on a classic inhibitory control Stroop task (Stroop, 1935). The competing processes account of performance is further supported by studies that have discovered that the link between dot comparison task performance and formal mathematics achievement can be accounted for by inhibitory control skills (Fuhs & McNeil, 2013; Gilmore et al., 2013).

Challenges to the dominant number sense theory of ANS processing are becoming increasingly frequent, however dot comparison tasks remain frequently reported in the literature as a measure of ANS acuity, without discussion of or control for visuospatial processing and inhibitory control skills.

The aim of this study was to investigate whether, in order to explain performance on dot comparison tasks, it is sufficient to consider *only* the visual cues contained in the task

stimuli¹. In other words, we ask whether the assumption that the ANS exists is required to account for participants' behaviour. To explore this question we calculated the extent to which we could explain performances on a standard dot comparison task with an extremely simple model where only visual cues influence judgements, and where the manner in which these visual cues influence judgements does not vary between trials. We then determined whether adding numerical information to this model significantly improved its predictive power.

Previous evidence has shown that performance on dot comparison tasks improves during adolescence (Halberda et al., 2012), and the relationship between dot comparison task performance and mathematics achievement may decline with age (Fazio, Bailey, Thompson & Siegler, 2014; Inglis et al., 2011). This raises the possibility that dot comparison tasks measure different cognitive skills at different stages of development, and so it is important to study individual differences in approaches to dot comparison tasks across development.

Similarly, evidence has shown that the method used to construct dot comparison stimuli significantly affects task performance (Clayton et al., 2010; Smets et al., 2015). Therefore, using stimuli created with different methods is important when exploring the role of visual cues in dot comparison task behaviour.

In sum, we had three main research questions. First, can performance on dot comparison tasks be explained using visual cues alone? Second, are there developmental differences with respect to this question? Finally, are there significant differences across tasks that use different stimuli construction methods?

Method

A total of 244 participants (124 children, mean age = 9.19, $SD = 1.25$; 120 adults, mean age = 22.86, $SD = 3.85$) completed a dot comparison task in three separate studies.

¹ The visual cues we chose to explore were average dot size and convex hull size. Although some studies also report controlling for additional variables such as cumulative surface area and density, these two factors are highly correlated with average dot size (Gebuis & Reynvoet, 2012), and so there would be no substantial benefit to examining these as separate predictors. Some methods for constructing dot arrays control only for cumulative surface area. Where we used stimuli originally constructed in this way, we nevertheless calculated dot size and convex hull size for these stimuli and used this in our analyses.

The studies were approved by the Loughborough University Ethics Approvals (Human Participants) Sub-Committee. Details of each study are described below.

Study 1

Participants

Participants were 51 adult students from Loughborough University (19 male, 32 female), with a mean age of 24.47 years ($SD = 4.50$), and 80 children aged 7-11 years (42 male, 38 female), with a mean age of 9.65 years ($SD = 1.27$). Participants were tested individually in a quiet room. Adults were given a £4 inconvenience allowance for their time, and children received game tokens as part of a Summer Scientist Week event at the University of Nottingham (www.summerscientist.org).

Task procedure

On each trial participants briefly viewed two arrays of dots presented simultaneously, side-by-side on a 15" laptop screen. Participants were required to select the more numerous array using the left and right keys marked on the keyboard. Instructions were read out loud and presented on screen to the participants as follows: "You are going to see two sets of dots. You have to decide which set has the most dots in it. If this side of the screen contains the most dots press the left key [text on left side of screen, stickered 'a' key pointed out by experimenter] / if this side of the screen contains the most dots press the right key [text on right side of screen, stickered 'l' key pointed out by experimenter]. Each trial began with a fixation point (600ms) followed by the presentation of the two arrays (600ms), and finally a grey screen with a white '?' presented in the centre until a response was given. There were 8 practice trials followed by 96 experimental trials. Trials were presented in a random order. The task took approximately 5 minutes to complete and breaks were given throughout. This task was part of a larger battery of numerical and cognitive processing tasks that are not reported here.

Stimuli

The stimuli were created using the Gebuis and Reynvoet (2011) protocol to control for visual cues. This created four image types. Image type 1 included pairs of arrays where

the more numerous array contained bigger dots and a bigger convex hull (fully congruent). Image type 2 included pairs of arrays where the more numerous array contained smaller dots and bigger convex hull (dot incongruent, convex hull congruent). Image type 3 included pairs of arrays where the more numerous array contained bigger dots and a smaller convex hull (dot congruent, convex hull incongruent). Image type 4 included pairs of arrays where the more numerous array contained smaller dots and a smaller convex hull (fully incongruent). The ratios between the number of dots in each array (calculated larger/smaller) ranged from 1.14 – 1.64 and the absolute numbers represented ranged from 22 – 36 dots. The array of dots on the left was coloured yellow and the array on the right was coloured blue, presented on a grey background.

Study 2

Participants

Participants were 57 adult students from Loughborough University (24 male, 33 female) with a mean age of 21.34 years ($SD = 2.35$). Participants were tested individually in a quiet room and were given a £3 inconvenience allowance for their time.

Task procedure

Participants completed a dot comparison task made up of two types of trials: stimuli created using the Gebuis and Reynvoet (2011) protocol, and stimuli created using the Panamath software. Trial presentation followed the same procedure described in Study 1. There were eight practice trials followed by a total of 312 experimental trials, which were divided into four blocks. Block one consisted of 96 Gebuis and Reynvoet trials and block two consisted of 60 Panamath trials. Both blocks of trials were presented twice. The order of the blocks was counterbalanced and trials within the blocks were presented in a random order. The task took approximately 10 minutes to complete and breaks were given throughout. This task was part of a larger battery of numerical and cognitive processing tasks that are not reported here.

Stimuli

The Gebuis and Reynvoet trials were identical to those used in Study 1. The Panamath trials were identical to those used by Libertus et al. (2012); on half of the trials the numerically larger set had a larger cumulative surface area, and on the other half of the trials the numerically larger set had a smaller cumulative surface area. The ratios between the number of dots in each array ranged from 1.06 – 2 and the absolute numbers represented ranged from 10 – 24 dots. As described below, we calculated the average dot size and convex hull size for the Panamath stimuli and used these, rather than cumulative surface area, in our analyses, to allow direct comparison across stimuli types.

Study 3

Participants

Participants were 44 children (21 male, 23 female) aged 7-9 years with a mean age of 8.36 years ($SD = 0.60$), and 12 adults (3 male, 9 female) with a mean age of 23.20 years ($SD = 4.04$). Children were tested in a quiet area of their school and were given a certificate for taking part. Adult participants were tested individually in a quiet room and were given £2 inconvenience allowance for their time.

Task procedure

The data collected were part of a study exploring the influence of set size on dot comparison performance (Clayton & Gilmore, 2014). The entire procedure for this task was identical to that described in Study 1, except there were 184 experimental trials. The task took less than 10 minutes to complete and breaks were given throughout.

Stimuli

Stimuli were created with the Gebuis and Reynvoet (2011) protocol. The arrays consisted of white dots on a black background. The ratios between the number of dots in each array ranged from 1.07– 1.9 and the absolute numbers represented ranged from 10 – 72 dots.

Analysis

For each of the stimuli used in the three studies, the average dot size and the convex-hull size of each array was calculated. The Graham Scan algorithm (Graham, 1972) was used to calculate the size of the convex hull. Average dot size was calculated by summing the number of coloured pixels in each array and dividing by the number of dots. Using these values, the ratio differences between the two arrays comprising each trial were calculated (largest/smallest numerosity) for convex-hull, average dot size and number. These ratios were then log transformed to ensure linearity so that equivalent ratios above and below 1 were comparable. Ratio characteristics of the stimuli are shown in Table 1.

Each participant's trial-by-trial accuracy scores were subjected to separate hierarchical logistic regressions, predicting accuracy for every trial with two steps: step one included dot-size ratio and convex-hull size ratio (visual cues), step two included numerical ratio. The first step represents a model where judgements for each trial are entirely based on the dot-size and convex hull ratios of that trial, and where the relative weights of these ratios does not vary between trials. The second step represents a model where behaviour is a function of these visual cue ratios as well as the trial's numerical ratio. If numerical processing was fundamental to performance on the dot comparison task we would expect the second model to be a substantial improvement over the first.

The median Nagelkerke pseudo- R^2 change from the addition of step two was recorded. Further to this, whether or not the addition of numerical information in step two of the regression significantly improved the model was recorded as binary data (either significant or non-significant). This analysis aimed to capture whether, for each participant individually, accuracy on the dot comparison trials was significantly predicted by numerical ratio after visual cue information was taken into account.

These data from all 244 participants were combined across studies 1, 2 and 3. Two sets of analyses were conducted involving different sets of trials. First, only trials created with the Gebuis and Reynvoet method were considered because previous research has demonstrated a non-significant relationship between individuals' performances on tasks created with different controls for visual cues (Clayton et al., 2015; Smets et al., 2015). For this analysis data from all 244 participants were combined across the three studies. A Mann-

Whitney U test was used to compare differences in adults' and children's pseudo- R^2 increase due to the inclusion of numerical ratio in the model. Pearson's chi-squared tests were then used to examine whether adults and children differed in their use of numerical information, as measured by whether or not the addition of numerical ratio in step two of the regression significantly improved the model, after visual cues were controlled for. Chi-squared tests were also used to examine whether there was any difference in use of numerical information across the three studies (Studies 1, 2 and 3).

Second, data from the 57 participants in Study 2 were examined to explore whether different protocols have any influence on adults' use of numerical information independent from visual cues. For this analysis the hierarchical logistic regression was performed twice for each participant, once with the Gebuis and Reynvoet trials, and once with the Panamath trials. A McNemar test was used to compare differences in whether numerical information significantly improved the regression model, over and above visual cues, between trials created with each protocol.

Results

Descriptive statistics, including mean accuracy scores and standard deviations are provided in Figure 1. The median pseudo- R^2 values for each step of the regression models are reported in Table 2, and the median standardised beta weights from the second stage of the model, (calculated as per Menard, 2011, formula 5) are reported in Table 3. For each study, the changes in pseudo- R^2 values, and the percentage of participants for whom numerical ratio significantly improved the regression model after controlling for visual cues, are presented in Table 4. Figure 2 shows that adults demonstrated larger increases in pseudo- R^2 values due to the addition of numerical information in the model at step two, when controlling for visual cue information in step one. This increase in pseudo- R^2 values for the adults ($Mdn = 0.055$) represented a significantly larger increase in comparison to the change in children's pseudo- R^2 values ($Mdn = 0.015$), $U = 3836.5$, $p < .001$. In line with this, a chi-squared test of independence showed a significant effect of age group on whether or not numerical information significantly improved the model when added in step two of the regression, $\chi^2(1,$

$N = 244$) = 41.06, $p < .001$, $\Phi = .41$. The addition of numerical information to the model explained significantly more variance in accuracy scores than visual cues alone for 71.7% of adults, and just 30.6% of children. This implies that for a majority of children (69.4%), and a large minority of adults (28.3%), accuracy on dot comparison trials could be accounted for without the need to include numerical ratio in the model.

This effect was consistent across multiple dot comparison studies for both adults and children. Chi-squared tests of independence showed no effect of study characteristics (e.g. amount of trials, range of numerical ratios presented) on whether numerical information significantly improved the regression model when added in step two, after controlling for visual cues at step one. The effects of study characteristics were non-significant for children, $\chi^2(1, N = 124) = 1.05$, $p = .306$, and adults $\chi^2(2, N = 120) = 5.267$, $p = .072$.

A final analysis was conducted with the data from Study 2 to explore the influence of the protocol used to construct dot array stimuli (i.e. Gebuis & Reynvoet; Panamath). A McNemar test demonstrated that the method of stimuli construction had a significant effect on whether or not adding numerical information significantly improved the fit of the regression model in step two, after accounting for visual cues, $p = .004$. Adult participants were more likely to use numerical information over and above visual cue information on trials created with the Gebuis and Reynvoet protocol (78.9%) than on trials created with the Panamath protocol (52.6%).

It is worth noting that when the steps of the regression model were reversed to include numerical ratio at step one, and visual cues (dot-size ratio and convex-hull ratio) at step two, pseudo- R^2 increased significantly for both adults ($Mdn = 0.341$) and children ($Mdn = 0.326$) at step two. There were no significant differences in pseudo R^2 changes across age groups ($U = 7131$, $p = .576$), but a chi-squared test of independence showed a significant effect of age group on whether or not visual cues significantly improved the fit of the model when added in step two of the regression, $\chi^2(1, N = 244) = 4.429$, $p = .035$, $\Phi = .135$. The addition of visual cue information to the model explained significantly more variance in accuracy scores than numerical information alone for 98.3% of adults and 92.7% of children. There was no effect of study characteristics on whether visual cue information significantly

improved the fit of the model when added in step two, after controlling for numerical information at step one, for children $\chi^2(1, N = 124) = 1.7, p = .191$, and adults $\chi^2(2, N = 120) = 2.752, p = .253$. Finally, a McNemar test demonstrated that stimuli generation method had no significant effect on whether or not visual cues accounted for significantly greater variance in participants' accuracy scores over and above numerical information, $p = .063$. Adult participants were as likely to use visual cues over numerical information on trials created with the Gebuis and Reynvoet protocol (100%), as on trials created with the Panamath protocol (91.2%).

Discussion

Nonsymbolic dot comparison tasks are assumed to measure ANS acuity, but few studies have explored the validity of this widely used task. Recent evidence has highlighted the significant influence of visual cue processing on dot comparison performance (Clayton & Gilmore, 2014; Fuhs & McNeil, 2013; Gebuis & Reynvoet, 2012; Gilmore et al., 2013; Smets et al., 2015; Szűcs et al., 2013), however there is a dearth of research investigating individual and developmental differences in the cognitive underpinnings of task performance. Here, we investigated whether or not we need the hypothesis that the ANS is involved in comparing nonsymbolic dot arrays. More specifically, we asked whether we can explain participants' behaviour on this task using a simple model that assumed that judgements are entirely driven by integrating information derived from visual cues, and that the manner of this integration does not vary between trials. We found that, for the majority of children in the study, adding numerical information over and above the information provided by visual cues did not explain significantly greater variance. For most adults, however, adding numerical information did improve the model. There were no significant differences in these findings across the three studies created using the same method of stimuli construction. These results raise the possibility that nonsymbolic comparison tasks are completed using different strategies by different age groups.

Our simple model assumed that each participant used visual cues in the same way on all the trial they completed. However, this is a gross oversimplification: for instance, research

suggests that participants are likely to switch their focus between different visual cues on each trial depending on the characteristics of both the individual trial and those of recent trials (Gebuis & Reynvoet, 2011; 2012). It is therefore likely that, with a more realistic model where the effect of visual cues is allowed to vary on a trial-by-trial basis, the assumption that numerical processing is taking place will be required for even fewer participants.

Consequently, our simple model has, if anything, overestimated participants' reliance on numerical information.

Our findings can help us to distinguish between the three prominent models of nonsymbolic comparison processing presented in our introduction: the dominant number sense model of the ANS (Barth et al., 2005; Dehaene, 1997; Feigenson et al. 2004), the visual cues hypothesis (Gebuis & Reynvoet, 2012), and the competing processes account (Clayton & Gilmore, 2015; Fuhs & McNeil, 2013; Gilmore et al., 2013; Nys & Content, 2012). First, the traditional ANS explanation of dot comparison performance, which suggests that ANS processing occurs independently of visual cues, cannot explain the present results. When the regression analysis was reversed, with numerical information included at step one, the addition of visual cues at step two significantly improved the model for almost every participant, suggesting that the vast majority of participants (>90%) did not complete the task based on numerical judgements alone. Second, the opposite hypothesis that comparison judgements can be made solely by weighing up visual cues holds up for 69.4% of children's data and just 28.3% of adults. In our over-simplified model where the relative salience of visual cues did not change on a trial-by-trial basis, including numerical information in the model explained significantly greater variance in performance than a model with visual cues alone. That is not to say that a more complex visual cue account where trial-by-trial changes in visual cue saliency are accounted for could not better explain the present results. Therefore, the visual cue model cannot be ruled out, but it would need expanding to explain the developmental differences in how visual cues are processed. In the current forms of the three accounts, the competing processes account is most consistent with the present results. This account accounts for individual and developmental differences in performance through the role of inhibition. Previous evidence suggests that individual differences in dot comparison

performance can be, at least partially, attributed to individual differences in inhibitory control skills (Gilmore et al., 2013; Cappelletti, Didino, Stoianov, & Zorzi, 2014). It could be that the developmental differences evidenced here are a result of improved inhibitory control development between children and adults (Luna et al., 2004; Morton, 2010). To clarify, adults who have superior inhibitory control skills may be more proficient and ignoring irrelevant visual features of the stimuli and focusing on numerosity, in comparison to children with poorer inhibitory control skills who may not be able to overcome the influence of these visual cues. Of course this hypothesis requires further investigation with children and adults in a study where inhibitory control ability is also measured.

Salti, Katzin, Katzin, Leibovich and Henik (2016) proposed that differences in saliency between visual and numerical cues might explain the strategies that participants use when comparing dot arrays. Specifically, Salti et al. suggested that when the differences between visual cues in a dot comparison trial are more salient than the differences between numerosities, it is plausible that participants would rely on the more salient dimension, visual cues, to guide their response. However, this hypothesis cannot explain the results of the present study for three reasons. Firstly, the Panamath stimuli had similar ratios across the three predictors (convex-hull ratio, dot-size ratio and numerical ratio; see Table 1), yet the standardised betas (see Table 3) showed the same patterns as the Gebuis and Reynvoet stimuli which had larger ratios for visual cues. Secondly, the salience of the ratios does not appear to be the critical factor in the regression analyses; Convex-hull ratios were smaller than dot-size ratios, yet the standardised betas in the regression were bigger for convex-hull. In contrast, Salti et al.'s (2016) proposal would predict dot size to be the most salient cue, and therefore the strongest predictor of dot comparison performance. Thirdly, adults and children completed the same dot comparison task trials, and so differences in cue saliency cannot explain the developmental effects found. In sum, although it is possible that differences in saliency between cues might bias participant's strategies to some extent, this is unlikely to be the only factor guiding participants' responses.

As well as the aforementioned theoretical implications, the present findings have methodological implications for the use of dot comparison tasks in research intending to

assess ANS acuity. First, and most crucially, our findings suggest that dot comparison tasks are not suitable as a measure of ANS acuity for all children. For almost 70% of children, numerical judgements did not explain significant extra variance in accuracy scores over and above that explained by visual cues, and therefore we cannot be sure that any numerical processing took place. This has serious implications for studies that have investigated the correlation between dot comparison performance and symbolic mathematics achievement. Conclusions about this relationship that are based on dot comparison performance as a measure of ANS acuity may be invalid. This is particularly important because a large proportion of the studies investigating the acuity of numerical representations that may underlie mathematical achievement have been conducted with school-aged children. In particular, some studies have demonstrated that a stronger association between dot comparison performance and mathematics ability is found with children rather than adults (Fazio, et al., 2014; Inglis et al., 2011). This has often been interpreted as a correlation between ANS acuity and mathematics achievement, however these results could also be caused by a mutual relationship with other cognitive skills, such as inhibitory control.

Second, one can conclude from this study that dot comparison tasks measure different cognitive constructs in adults in comparison to 7-11 year old children. Adults were significantly more likely than children to use numerical information over and above visual cue information when comparing dot arrays. From this, future research should not assume the same underlying processes contribute to accuracy scores for different developmental groups. Dot comparison tasks appear to be better measures of ANS acuity for adults, however, this group was far from homogeneous, and still 28% of our adult participants did not appear to have used numerosity information over and above visual characteristics.

Our results were consistent across three studies including dot comparison tasks that varied in the range of numerosities represented and the number of trials completed by the participants. For a subset of 57 adult participants, the influence of the protocol for creating the dot array stimuli was also analysed. Participants were more likely to be influenced by numerical information independently from visual cue information on trials created with the Gebuis and Reynvoet protocol (designed to multiple visual cues simultaneously, including

convex hull), than on trials created with the Panamath protocol (designed to manipulate single visual cues, excluding convex hull). A possible explanation for this difference is that due to the lack of systematic control over convex hull size in the Panamath dot arrays, participants were more able to make effective judgements on numerosity based on visual cues alone (see Clayton et al., 2015, for evidence of the confound between convex hull and numerosity in Panamath stimuli). Nevertheless, findings from the three studies that used the Gebuis and Reynvoet stimuli generation protocol appear robust despite several methodological distinctions between tasks. Future research will be important to assess whether results are consistent across other versions of nonsymbolic comparison tasks. Moreover, the results of this study are limited to dot comparison tasks; further research could use this method to investigate whether numerical information is predictive of performance on other tasks designed to measure ANS acuity, such as nonsymbolic estimation and nonsymbolic arithmetic tasks.

In sum, we have shown that for a significant number of participants, there is no need to hypothesise that the ANS exists to explain their performance on dot comparison tasks. To be precise, for the majority of children and some adults, numerical processing did not explain significant additional variance in dot comparison task performance over and above visual cue processing. This finding has implications for research that has shown a correlation between nonsymbolic dot comparison performance and symbolic mathematical ability, as it appears likely that this relationship is not caused by the assumed mutual relationship with the ANS, especially for children. If psychologists are to continue to use nonsymbolic comparison tasks as a measure of ANS acuity in their research, analyses to evaluate whether participants are processing numerical information should be completed before conclusions are drawn, and other domain-general cognitive skills such as inhibition and visuospatial skills should be considered. This result, combined with recent research highlighting substantial issues concerning the reliability and validity of different dot comparison task methodologies (Inglis & Gilmore, 2014; Smets et al., 2015), suggests great care is required when using dot comparison tasks to investigate numerical cognition.

References

- Barth, H., La Mont, K., Lipton, J., & Spelke, E. S. (2005). Abstract number and arithmetic in preschool children. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(39), 14116–14121. doi:10.1073/pnas.0505512102
- Barth, H., La Mont, K., Lipton, J., Dehaene, S., Kanwisher, N., & Spelke, E. (2006). Non-symbolic arithmetic in adults and young children. *Cognition*, *98*(3), 199–222. doi:10.1016/j.cognition.2004.09.011
- Cappelletti, M., Didino, D., Stoianov, I., & Zorzi, M. (2014). Number skills are maintained in healthy ageing. *Cognitive Psychology*, *69*, 25 – 45. doi:10.1016/j.cogpsych.2013.11.004
- Chen, Q., & Li, J. (2014). Association between individual differences in non-symbolic number acuity and math performance: A meta-analysis. *Acta Psychologica*, *148*, 163-172. doi:10.1016/j.actpsy.2014.01.016
- Clayton, S., & Gilmore, C. (2015). Inhibition in dot comparison tasks. *ZDM Mathematics Education*. Advance online publication. doi:10.1007/s11858-014-0655-2
- 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. doi:10.1016/j.actpsy.2015.09.007
- Dehaene, S. (1997). *The number sense*. Oxford: Oxford University Press.
- De Smedt, B., Noël, M.-P., Gilmore, C., & Ansari, D. (2013). How do symbolic and non-symbolic numerical magnitude processing skills relate to individual differences in children's mathematical skills? A review of evidence from brain and behavior. *Trends in Neuroscience and Education*, *2*, 48-55. doi:10.1016/j.tine.2013.06.001
- 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. doi:10.1016/j.jecp.2014.01.013
- Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences*, *8*(7), 307–314. doi: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*, 136–148. doi: 10.1111/desc.12013
- Gebuis, T., Kadosh, R. C., de Haan, E., & Henik, A. (2009). Automatic quantity processing in 5-year olds and adults. *Cognitive Processing*, *10*(2), 133–142. doi:10.1007/s10339-008-0219-x
- Gebuis, T., & Reynvoet, B. (2011). Generating nonsymbolic number stimuli. *Behavior Research Methods*, *43*, 981–986. doi:10.3758/s13428-011-0097-5
- Gebuis, T., & Reynvoet, B. (2012). The interplay between nonsymbolic number and its continuous visual properties. *Journal of Experimental Psychology: General*, *141*(4), 642–648. doi:10.1037/a0026218
- 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. doi:10.1371/journal.pone.0067374
- Graham, R. L. (1972). An efficient algorithm for determining the convex hull of a finite planar set. *Information Processing Letters* *1*, 132-133. doi:10.1016/0020-0190(72)90045-2
- 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. doi:10.1073/pnas.1200196109
- 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. doi:10.1016/j.jecp.2008.04.001
- Hurewitz, F., Gelman, R., & Schnitzer, B. (2006). Sometimes area counts more than number. *Proceedings of the National Academy of Sciences*, *103*(51), 19599–19604. doi: 10.1073/pnas.0609485103
- Inglis, M., & Gilmore, C. (2014). Indexing the Approximate Number System. *Acta*

- Psychologica*, 145, 147-155. doi:10.1016/j.actpsy.2013.11.009
- Inglis, M., Attridge, N., Batchelor, S., & Gilmore, C. (2011). Non-verbal number acuity correlates with symbolic mathematics achievement: But only in children. *Psychonomic Bulletin & Review*, 18(6), 1222–1229. doi:10.3758/s13423-011-0154-1
- 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. doi:10.1017/S0140525X16000960 [Epub ahead of print]
- Luna, B., Garver, K. E., Urban, T. A., Lazar, N. A., & Sweeney, J. A. (2004). Maturation of cognitive processes from late childhood to adulthood. *Child development*, 75(5), 1357-1372. doi:10.1111/j.1467-8624.2004.00745.x
- Menard, S. (2011). Standards for Standardized Logistic Regression Coefficients. *Social Forces*, 89(4), 1409-1428. doi:10.1093/sf/89.4.1409
- Morton, J. B. (2010). Understanding genetic, neurophysiological, and experiential influences on the development of executive functioning: the need for developmental models. *Wiley Interdisciplinary Reviews: Cognitive Science* 1(5), 709–723. doi:10.1002/wcs.87
- 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. doi:10.1080/17470218.2011.619661
- 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. doi:10.3758/s13428-016-0772-7
- 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*. doi:10.1111/desc.12372
- Smets, K., Sasanguie, D., Szűcs, D., & Reynvoet, B. (2015). The effect of different methods to construct non-symbolic stimuli in numerosity estimation and comparison. *Journal of Cognitive Psychology*, 27(3), 310–325. doi:10.1080/20445911.2014.996568
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental*

Psychology: General, 18(6), 643–662. doi:10.1037/h0054651

Szűcs, D., Nobes, A., Devine, A., Gabriel, F. C., & Gebuis, T. (2015). Visual stimulus parameters seriously compromise the measurement of approximate number system acuity and comparative effects between adults and children. *Frontiers in Psychology*, 4, 1–12. doi:10.3389/fpsyg.2013.00444

Figure 1.

Mean accuracy scores for Gebuis and Reynvoet trials (G&R) and Panamath trials split by image congruency type, labeled as follows: Fully cong = convex-hull congruent, dot-size congruent; CH Cong = convex-hull congruent, dot-size incongruent; Dot Cong = convex-hull incongruent, dot-size congruent; Fully Incong = convex-hull incongruent, dot-size incongruent. Error bars represent standard deviations.

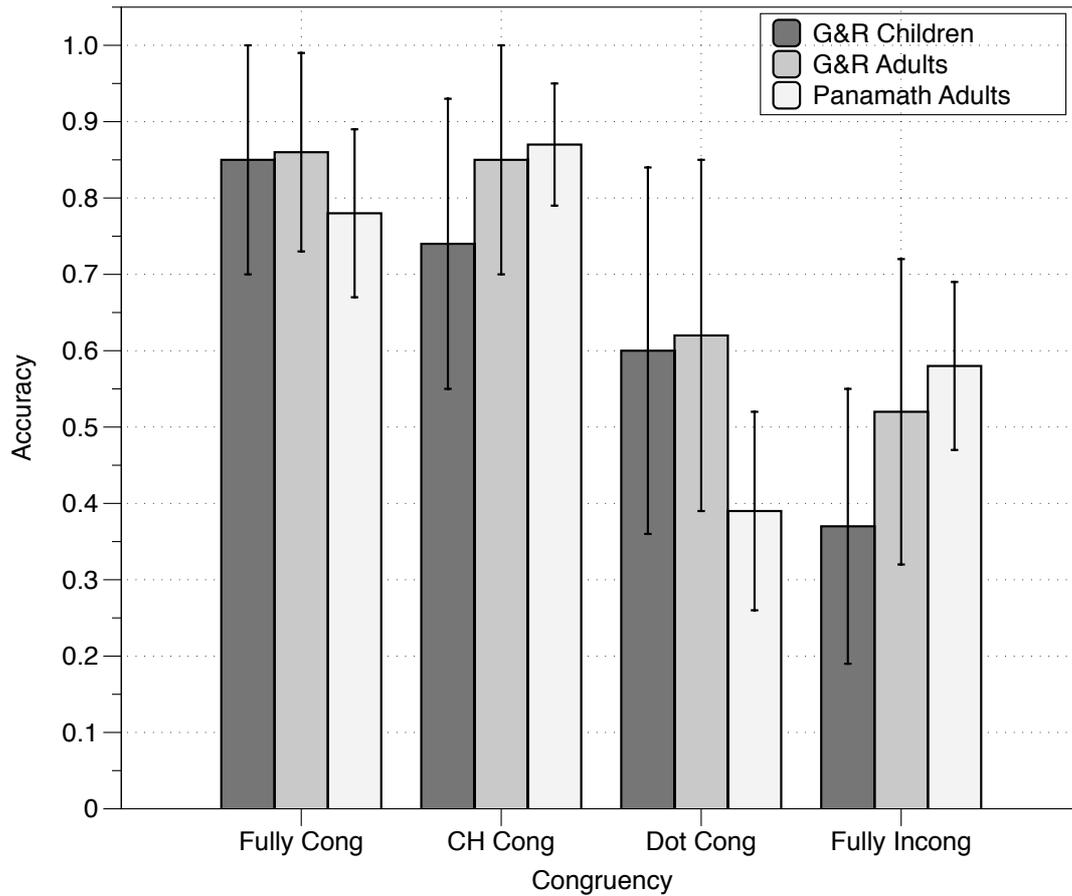


Figure 2.

Change in pseudo- R^2 values when numerical ratio was added to regression models individually predicting accuracy scores for children (top panel) and adults (bottom panel).

Data from the Panamath trials were not included here.

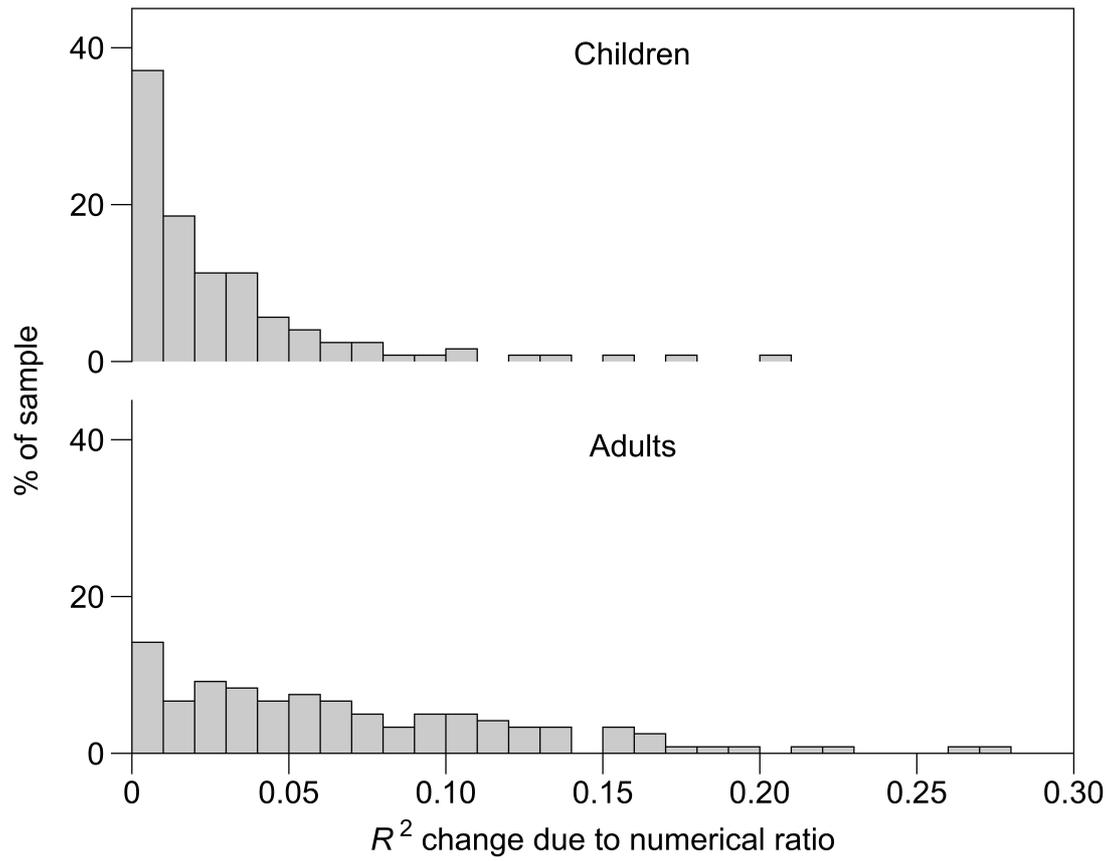


Table 1.

Numerosity and visual cue information for the stimuli in each of the three studies, including the mean and the standard deviations of the ratios between the to-be-compared arrays in terms of numerosity, average dot size and convex hull size. All ratios are calculated larger value / smaller value for each cue to allow for comparison and log transformed. (G&R = Gebuis & Reynvoet).

	Numerosity ratio		Dot size ratio		Convex hull ratio	
	M	SD	M	SD	M	SD
Study 1 – G&R	0.13	0.06	0.63	0.14	0.24	0.07
Study 2 – G&R	0.13	0.06	0.63	0.14	0.24	0.07
Study 2 – Panamath	0.10	0.08	0.13	0.15	0.07	0.06
Study 3 – G&R	0.13	0.08	0.61	0.15	0.23	0.08

Table 2.

The median pseudo- R^2 values for step one and step two of the regression models.

		Gebuis and Reynvoet stimuli				Panamath stimuli	
		Children		Adults		Adults	
		Step 1	Step 2	Step 1	Step 2	Step 1	Step 2
Median pseudo- R^2	Study 1	.347	.388	.344	.385	-	-
	Study 2	-	-	.399	.449	.296	.343
	Study 3	.324	.359	.205	.368	-	-
	Overall	.337	.365	.357	.414	.296	.343

Table 3.

The median standardised beta weights when convex-hull ratio, dot-size ratio, and numerical were all added to the regression models (step two). (G&R = Gebuis & Reynvoet).

		G&R stimuli		Panamath stimuli
		Children	Adults	Adults
Convex Hull	Study 1	.337	.361	-
	Study 2	-	.494	.306
	Study 3	.399	.298	-
	Overall	.355	.403	.306
Dot size	Study 1	.178	.138	-
	Study 2	-	.040	-.086
	Study 3	.122	-.105	-
	Overall	.168	.053	-.086
Numerical	Study 1	.140	.249	-
	Study 2	-	.228	.292
	Study 3	.120	.317	-
	Overall	.128	.247	.292

Table 4.

The median pseudo- R^2 change when numerical ratio was added to the regression models and the percentage of participants for whom adding numerical ratio to the regression model accounted for significantly greater variance in accuracy scores, after controlling for visual cues, across all three studies. Data from the Panamath trials were not included here.

	Children		Adults	
	Median R^2 change	% sig	Median R^2 change	% sig
Study 1	0.017	27.5%	0.065	60.8%
Study 2	-	-	0.049	78.9%
Study 3	0.013	36.4%	0.095	83.3%
Overall	0.015	30.6%	0.055	71.7%