Journal of Cognitive Psychology

ISSN: 2044-5911 (Print) 2044-592X (Online) Journal homepage: http://www.tandfonline.com/loi/pecp21

The effect of different methods to construct non-symbolic stimuli in numerosity estimation and comparison

Karolien Smets, Delphine Sasanguie, Dénes Szücs & Bert Reynvoet

To cite this article: Karolien Smets, Delphine Sasanguie, Dénes Szücs & Bert Reynvoet (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

To link to this article: https://doi.org/10.1080/20445911.2014.996568

Published online: 03 Jan 2015.

Journal of Cognitive Psychology, 2015
Vol. 27, No. 3, 310–325, http://dx.doi.org/10.1080/20445911.2014.996568

The effect of different methods to construct non-symbolic stimuli in numerosity estimation and comparison

Karolien Smets[1], Delphine Sasanguie[1,2], Dénes Szücs[3], and Bert Reynvoet[1,2]

[1]Laboratory of Experimental Psychology, KU Leuven, 3000 Leuven, Belgium
[2]Faculty of Psychology and Educational Sciences Kulak, 8500 Kortrijk, Belgium
[3]Department of Psychology, Centre for Neuroscience in Education, University of Cambridge, Cambridge CB2 3EB, UK

(*Received 24 July 2014; accepted 3 December 2014*)

Numerosity estimation and comparison tasks are often used to measure the acuity of the approximate number system (ANS), a mechanism which allows extracting numerosity from an array of dots independently from several visual cues (e.g. area extended by the dots). This idea is supported by studies showing that numerosity can be processed while these visual cues are controlled for. Different methods to construct dot arrays while controlling their visual cues have been proposed in the past. In this paper, these methods were contrasted in an estimation and a comparison task. The way of constructing the dot arrays had little impact on estimation. In contrast, in the comparison task, participants' performance was significantly influenced by the method that was used to construct the arrays of dots, indicating better performance when the visual cues of the dot arrays (partly) co-varied with numerosity. The present study therefore shows that estimates of ANS acuity derived from comparison tasks are inconsistent and dependent on how the stimuli are constructed. This makes it difficult to compare studies which utilised different methods to construct the dot arrays in numerosity comparison tasks. In addition, these results question the currently held view of the ANS as capable of robustly extracting numerosity independently from visual cues.

Keywords: ANS acuity; Comparison; Estimation; Validity; Visual cue control.

Reasoning with non-symbolic numerosities refers to an everyday skill that humans share with non-human species (e.g. Agrillo, Piffer, & Bisazza, 2011; Meck & Church, 1983; Pahl, Si, & Zhang, 2013). It is suggested to be rooted in a universal, basic system, named the "approximate number system" (ANS; Brannon, 2006; Cordes, Gelman, Gallistel, & Whalen, 2001; Feigenson, Dehaene, & Spelke, 2004; Libertus & Brannon, 2010). By means of this ANS, humans and other species are assumed to be able to estimate the number of, for instance an array of dots or to discriminate between two non-symbolic numerosities. The ANS is presumed to represent non-symbolic numerosities approximately on a mental number line (Dehaene, 1997; Restle, 1970). Additionally, the ANS adheres to the Weber–Fechner law (Fechner, 1860) which states that to obtain a constant level of discrimination performance, an increasing larger difference between the two

Correspondence should be addressed to Karolien Smets, Experimental Psychology, KU Leuven, Tiensestraat 102, Post Box 3711, 3000 Leuven, Belgium. E-mail: karolien.smets@ppw.kuleuven.be
No potential conflict of interest was reported by the authors.
The research was supported by the Research Fund KU Leuven.

numerosities that have to be discerned is necessary when the numerosities increase.

The acuity of the ANS can be measured with a wide variety of different tasks (e.g. Gebuis & van der Smagt, 2011; Gilmore, Attridge, & Inglis, 2011; Price, Palmer, Battista, & Ansari, 2012; Sasanguie, Göbel, Moll, Smets, & Reynvoet, 2012). For instance, in the numerosity estimation task, participants are presented with an array of dots and are subsequently asked to estimate the number of dots present (Mejias, Grégoire, & Noël, 2012). One of the measures that reflect ANS acuity in studies using the estimation task is the precision of the estimates (i.e. the absolute error [AE] score; Mejias, Grégoire et al., 2012; Mejias & Schiltz, 2013). Studies using numerosity estimation consistently show that participants make more errors and are more variable in their responses when estimating a larger numerosity (e.g. 50 dots) compared with a smaller numerosity (e.g. 10 dots) (Gebuis & Reynvoet, 2012; Izard & Dehaene, 2008): There is an increase in mean estimates and standard deviations (*SDs*) of the participants' estimates with increasing target numerosity (Castronovo & Göbel, 2012; Mejias, Grégoire et al., 2012; Mejias & Schiltz, 2013).

In other studies, an estimate of ANS acuity is derived by means of a comparison task. In this task, participants are instructed to indicate the larger of two presented numerosities (e.g. Piazza et al., 2010; Price et al., 2012). Typically, a ratio effect in accuracy and reaction time is found: participants will be less accurate and slower when the ratio between the two numerosities that have to be discriminated is closer to one (Barth, Kanwisher, & Spelke, 2003; Defever, Reynvoet, & Gebuis, 2013; Halberda & Feigenson, 2008; Sasanguie et al., 2012). In the comparison task, the acuity of the ANS can, on the one hand, be expressed as an individual Weber fraction (w) for each participant (Halberda, Mazzocco, & Feigenson, 2008; Piazza et al., 2010), or on the other hand as the average accuracy, two measures that are highly correlated (Gilmore et al., 2011; Sasanguie et al., 2012).

One essential difference among the numerous studies investigating ANS acuity that has not received much attention in previous research is the effect of different methods used to construct the dot arrays. In daily life, numerosity and visual cues are highly correlated. For instance, if one compares two oranges to a more numerous set of 10 oranges in reality, the visual cue "area extended" (i.e. the convex hull or the smallest contour that can be drawn around the oranges) will also be larger in the more numerous set of oranges. In order to study numerosity processing while avoiding the confound of visual cues in the comparison and estimation task but also in other tasks (e.g. in the same-different task; Smets, Gebuis, & Reynvoet, 2013), researchers manipulate or control the visual cues of dot arrays, attempting to make them uninformative about numerosity (e.g. Gebuis & Reynvoet, 2011a; Gilmore et al., 2013; Pica, Lemer, Izard, & Dehaene, 2004; Sasanguie, Defever, Van den Bussche, & Reynvoet, 2011; Smets et al., 2013; Smets, Gebuis, Defever, & Reynvoet, 2014; Szücs, Nobes, Devine, Gabriel, & Gebuis, 2013).

Two methods have been used frequently to construct these dot arrays and control their visual cues in numerosity processing studies (but see also the more recent approach of Lyons, Price, Vaessen, Blomert, & Ansari, 2014). By means of a first method, developed by Dehaene, Izard, and Piazza (2005), dot arrays are created in which the visual cue "surface" (the sum of the different dot surfaces in one array) is kept constant in half of the trials and the visual cue "diameter of the dots" is kept constant in the other half of the trials (script available on http://www.unicog.org). It is assumed that this way participants are discouraged to base their judgement on the visual cues of the stimuli because if they do, their performance will be negatively affected considering no single visual cue is predictive for numerosity throughout all trials. The Matlab script to generate dot arrays this way (or related variants of this method) is a common approach in the literature and is viewed as a valid manner to verify that visual cues did not impact numerical performance. This technique has been used in several studies investigating ANS acuity or number sense (e.g. Halberda et al., 2008; Halberda, Ly, Wilmer, Naiman, & Gemine, 2012; Piazza, Pica, Izard, Spelke, & Dehaene, 2013; Revkin, Piazza, Izard, Cohen, & Dehaene, 2008; Sasanguie et al., 2011), for instance to relate basic number sense to higher-level mathematical achievement. Halberda et al. (2012; but see also Halberda et al., 2008) presented participants with either "dot-size controlled" or "area controlled" (described as the total number of pixels a dot array occupies) numerosity comparison stimuli in their massive Internet-based sample. These authors subsequently related the obtained number sense estimate to mathematical performance and concluded that individual differences in the precision of this number sense affect mathematics achievement. However, it should be noted that several researchers failed to find such a relation (e.g. Brankaer, Ghesquière, & De Smedt, 2014; Sasanguie et al., 2011).

Recently however, some authors (Gebuis & Reynvoet, 2011a, 2012; Szücs et al., 2013) warned that creating dot arrays and controlling their visual cues according to the method of Dehaene et al. (2005) or related techniques may not be sufficient. They did so by noting that participants may switch their focus between different visual cues or integrate multiple visual cues on a certain trial, meaning that participants can still benefit from the visual cues if the stimuli are created according to the method of Dehaene et al. (2005). To overcome this issue, Gebuis and Reynvoet (2011a) presented an alternative method to generate dot arrays in which the visual cues are controlled more stringently by manipulating several visual cues (script available on http://www.titiagebuis.eu/Materials.html). The latter way of controlling visual cues has several advantages (Gebuis & Reynvoet, 2011a, 2011b, 2012; Szücs et al., 2013). First, control of the relevant visual cues is more stringent because *more* visual cues are controlled for. Namely, four visual cues are taken into account: “area extended by the dots” (convex hull or smallest contour that can be drawn around the dots), “total surface” (the sum of the different dot surfaces within one array), “diameter of the dots” and “density” (surface divided by area extended or convex hull). Additionally, the script of Gebuis and Reynvoet (2011a) provides the concrete values for each of the visual cues for every dot array, which makes it possible to perform regression analyses to verify that there is in fact no relationship between the numerosities and their corresponding visual cues. A second advantage of creating dot arrays with the method of Gebuis and Reynvoet (2011a) is that no single visual cue is informative about numerosity across trials and that there is consequently no relationship between the difference in visual cues (e.g. area of the larger numerosity minus area of the smaller numerosity) and the associated difference in numerosity (Gebuis & Reynvoet, 2011a). Several researchers have adopted this method recently to create dot stimuli and to control their visual cues (e.g. Inglis & Gilmore, 2014; Katz & Knops, 2014; Linsen, Verschaffel, Reynvoet, & De Smedt, 2014; Smets et al., 2014; Szücs et al., 2013), mostly leading to an apparent lower performance than is usually obtained when generating dot arrays according to the method of Dehaene et al. (2005).

For instance, Szücs et al. (2013) used this more stringent type of visual cue control developed by Gebuis and Reynvoet (2011a) and obtained Weber fractions that were substantially larger than those in other studies in which a different manner of constructing the dot arrays was employed (e.g. Piazza et al., 2010). This suggests that the differential control of visual cues might have influenced participants’ performance. These discrepant Weber fractions in different studies investigating the ANS by means of a comparison task are problematic for several reasons. First, from the observation that performance is higher (and the Weber fraction is thus lower) in studies in which the visual cues of the dot arrays are controlled less strictly, we can conclude that participants, when possible, still rely on these visual cues to judge numerosities despite the use of visual cue controls. Second, if different Weber fractions are obtained when distinct methods to control the visual cues of dot arrays are used, this may lead ultimately to unstable and inconsistent estimates of ANS acuity. As a consequence, results may not be comparable across studies using different methods to construct dot arrays (i.e. the construct validity of the task at hand is questioned). This may be problematic as several researchers implicitly assume that at least Weber fractions are objective measures of ANS acuity and can be compared across studies with distinct methodological characteristics. These researchers attempt to describe the developmental trajectory of ANS acuity by providing a graph in which estimates of ANS acuity from different studies are combined (e.g. Halberda & Feigenson, 2008; Piazza, Izard, Pinel, Le Bihan, & Dehaene, 2004). Considering that the ANS may function as a core system for the support of higher-level mathematics (Feigenson et al., 2004), this developmental trajectory can potentially be used as a screening instrument to detect children who are at risk of developing mathematical difficulties. However, if estimates of ANS acuity obtained in studies with differing methodologies are incomparable, this developmental trajectory is currently poorly described and cannot be applied as a reliable screening instrument.

In the current study, we compared the influence of different ways of constructing dot arrays in a numerosity estimation task (Experiment 1) and in a numerosity comparison task (Experiment 2). In both experiments, the following three conditions were conducted. In a first condition, the dot arrays were created using the script of Dehaene et al. (2005) which keeps surface constant in half of the trials and diameter in the other half of the trials. In a second condition, the stimuli were generated according to the method of Gebuis and Reynvoet (2011a), which controls several visual parameters in such a way that they are uninformative about numerosity across trials. Finally, we included a third condition in which all stimuli consisted of numerosities with congruent visual cues by virtue of two reasons. First, we wanted to additionally explore the effects of *not* strictly controlling the visual cues in both numerosity estimation and comparison. A second reason for including this congruent condition was to enable more nuanced results with respect to the expected differences between the other two methods. More specifically, we wanted to explore to what extent both conditions in which visual cues are assumed to be controlled, differ from a condition where this is not the case. We specifically assessed potential differences between the three conditions concerning their estimates of ANS acuity. Additionally, correlations between the three conditions were computed. These correlation analyses were expected to provide indisputable evidence with regard to the comparability issue of the different conditions. More concrete, if an estimate of ANS acuity can be derived in a consistent manner and distinct methodologies do not strongly influence these estimates, performances in the three conditions are expected to be significantly correlated. To the best of our knowledge, the current study is the first in which the effects of different ways of generating dot arrays in the estimation or comparison task have been examined directly in adult participants.

EXPERIMENT 1: ESTIMATION TASK

Method

Participants. Forty-one adult participants took part in Experiment 1 (mean age = 23.22 years, SD = 4.68; 28 females). They were paid 8 Euros for their participation. The experiment was approved by the Ethical Committee of the Faculty of Psychology and Educational Sciences of the University of Leuven. All participants gave written informed consent.

Stimuli and procedure. Stimuli were presented on a 17-inch colour screen. Displaying of the stimuli and recording of the data were done with Matlab 7.1, using the Psychophysics Toolbox. Stimuli consisted of white dots on a black background. Participants were presented with three estimation conditions, differing in the method of constructing the dot arrays. Example stimuli for each of these estimation conditions and each target (see later for concrete information on the target numerosities) are presented in Figure 1. In a first condition, dot arrays where created with the script of Dehaene et al. (2005), in which surface was kept constant in half of the trials while diameter of the dots co-varied with numerosity (Figure 2a). In the other half of the trials, diameter was kept constant while surface co-varied with numerosity (Figure 2b). Figure 2 illustrates the hypothetical values of the visual cues “surface” and “dot diameter” as the veritable ones cannot be directly derived from the script developed by Dehaene et al. (2005). Because only one visual cue is kept constant in a single trial while the other cue co-varies with numerosity, we refer to this condition as the “single sensory control condition”. In a second condition, the script of Gebuis and Reynvoet (2011a) was used to construct the dot arrays. This script ensures that visual cue information is uninformative about numerosity by manipulating four visual cues: (1) area extended by the dots (convex hull), (2) surface (the sum of the individual dot surfaces in one array), (3) dot diameter and (4) density (surface divided by area). We refer to this condition as the “multiple sensory control condition”, because multiple visual cues are controlled for. The concrete and veritable average values over all trials for each visual cue for this method are illustrated in Figure 3. Post-hoc regression analyses verified that the different visual cues were uninformative about numerosity (all R^2 values < .10). Finally, in a third condition, which we refer to as the “congruent condition”, the visual cues completely co-varied with numerosity: larger numerosities were accompanied by a larger area extended and surface, but a smaller diameter. We opted for a connection between larger numerosities and smaller dot sizes because previous research indicated that participants associate smaller dot sizes with larger numerosities (Gebuis & Reynvoet, 2012). This congruent condition is more ecologically valid, considering the strong relation between numerosities and visual cues in everyday life. In all conditions, the numerosities that needed to be estimated were three target numerosities differing with a 2:3 ratio (12, 18 and 27). Additionally, we included some filler numbers (10, 14, 16, 21, 25 and 29) to prevent participants from noticing that only three

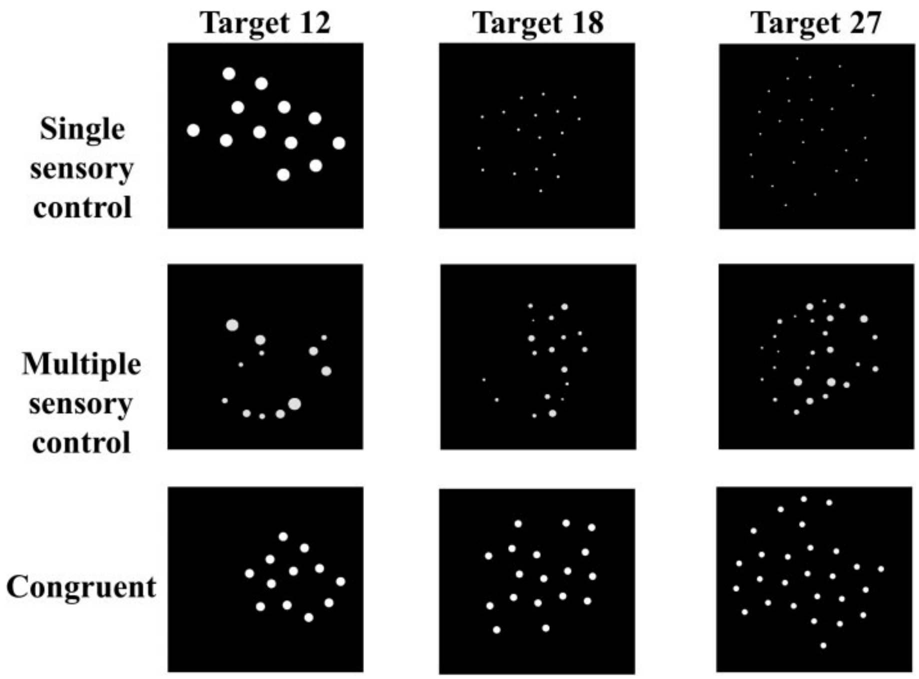

Figure 1. Examples of the three target stimuli for the single sensory control condition (example in which surface is kept constant, constructed with Dehaene et al., 2005), the multiple sensory control condition (constructed with Gebuis & Reynvoet, 2011a) and the congruent condition.

numerosities were used throughout the experiment (for a similar procedure, see Gebuis & Reynvoet, 2013).

In each trial of the three conditions, participants were presented with a dot array for 300 ms, after which a question mark appeared on the screen. Participants were instructed to estimate the number of dots by responding with the number keypad of the keyboard when the question mark appeared. After a response was given, a blank screen was displayed for 1250 to 1500 ms (for a similar procedure, see Gebuis & Reynvoet, 2012). Each target number was repeated 40 times, resulting in 120 target trials (= three targets × 40) and each filler number was repeated 12 times, resulting in 72 filler trials (= six fillers × 12). This way, each condition consisted of 192 trials in total (120 target trials + 72 filler trials), but only the 120 target trials were included in the analyses. There was a break in the middle of each of the three conditions. Participants performed all three conditions (single sensory control, multiple sensory control and congruent) following a Latin square distribution and could take additional breaks in between the conditions.

Data analyses. Outliers more than two *SDs* above or below the participant's own mean response for each target were excluded from the analyses (in total 4.32% of the trials; see also Gebuis and Reynvoet, 2012). We verified whether there was an effect of the order in which the participants performed the three conditions. Since no significant effect could be detected, order was not taken into account in the following analyses, $F(2,38) = .71$, $p = .50$, $\eta_p^2 = .04$.

Next, we checked for scalar variability which is expected in accordance to the Weber-Fechner law. Therefore, we performed regression analyses on the mean estimates in a first analysis and on the *SDs* in a second analysis with target (three levels: 12, 18 and 27) as predictor. In a third regression analysis, the coefficients of variation (COVs = *SD*/mean estimate; an indicator of variability of the estimates) were calculated and included as dependent variable, again with target as the predictor (for a similar procedure, see Mejias, Grégoire et al., 2012; Mejias & Schiltz, 2013). These three regression analyses were done separately for the three conditions: single sensory control (method of Dehaene et al., 2005), multiple sensory control (method of Gebuis & Reynvoet, 2011a) and congruent.

Subsequently, we calculated a measure that indicates the imprecision of the participants' estimates or the error rate by computing participants'

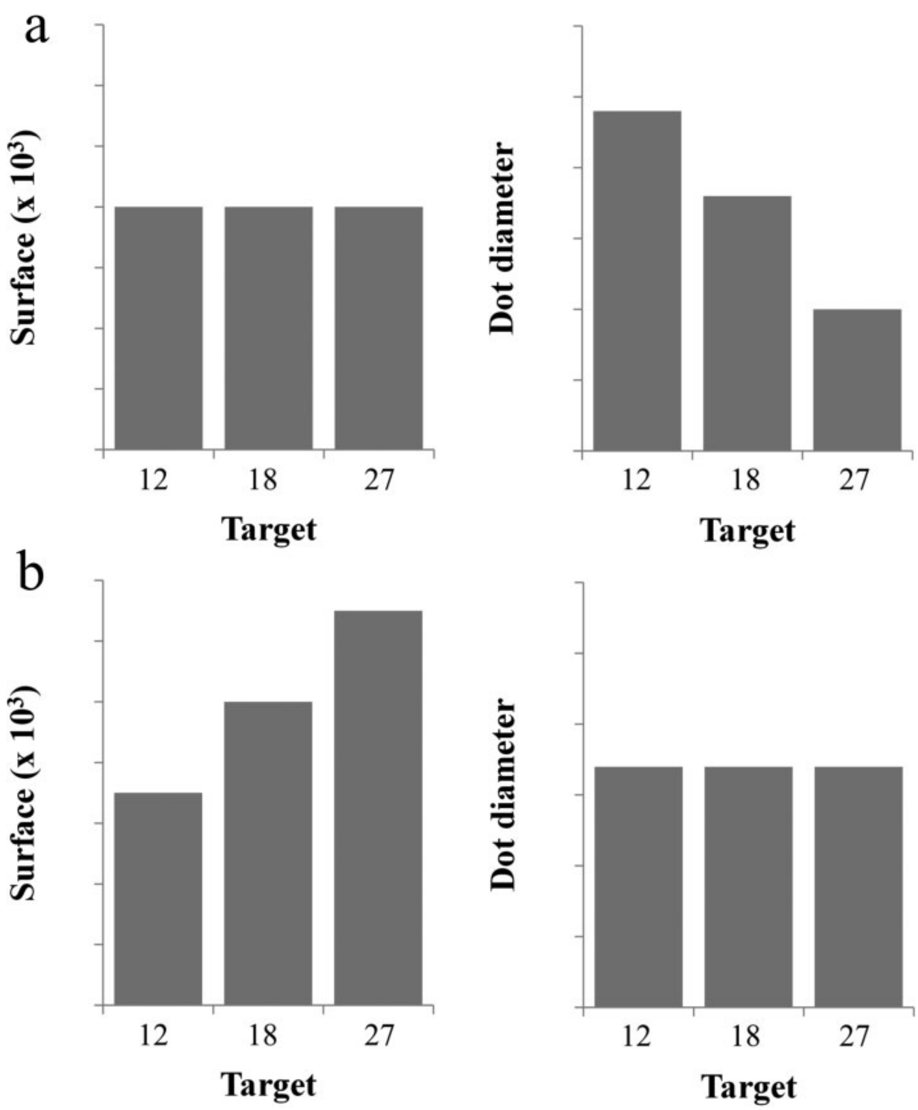

Figure 2. Clarification of the rationale behind the method of Dehaene et al. (2005). The upper images (a) hypothetically represent one half of the trials in which surface is kept constant, while diameter varies with numerosity. The bottom images (b) represent the other half of the trials with diameter constant, while surface varies with numerosity.

AE scores as a measure of ANS acuity. The AE score is calculated by means of the following formula: $AE = \Sigma|x_i - T|/n$, where x_i is the estimate provided by the participant on a given trial i, T is the actual target numerosity that is presented on that trial i and n is the number of trials each participant performed (Mejias, Grégoire et al., 2012; Schmidt & Lee, 2011). We opted to compute the AE score instead of calculating the mean error score (ER = [participant's mean response − target numerosity]/target numerosity) as was, for instance, done in the study of Crollen, Castronovo, and Seron (2011), because the AE score provides a general measure of imprecision that is independent of the direction of the difference between the participant's response and the actual target numerosity (see also Mejias, Grégoire et al., 2012; Mejias & Schiltz, 2013 for the use of the AE score as a measure in the estimation task). A repeated measures analysis with the AE score as the dependent variable and both target (three levels: 12, 18 and 27) and condition (three levels: single sensory control, multiple sensory control and congruent) as between-subjects variables was performed. We corrected the p values with the Greenhouse–Geisser method (pGG) when the assumption of sphericity was violated in the analyses. Finally, given the focus of our study on the comparability of the different conditions, we

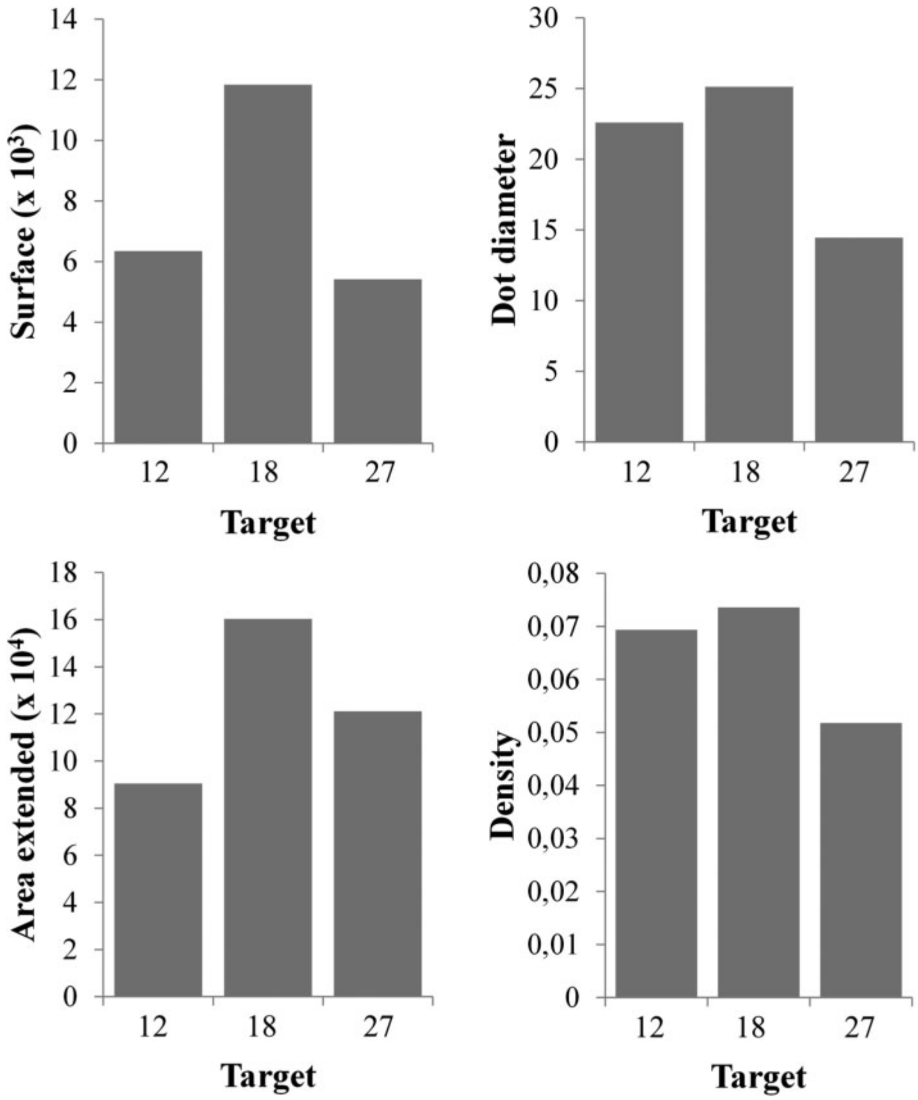

Figure 3. Calculated values across all trials for each of the visual cues for each target when dot arrays are created with the script of Gebuis and Reynvoet (2011a).

performed correlation analyses between the AE scores in the three conditions.

Results

The results first indicated that scalar variability was present in all three conditions (Table 1): the mean estimates increased with increasing target in all conditions, all R^2s > .41, all βs > .85, all *t*s > 9.15 and all *p*s < .001, as well as the *SDs*, all R^2s > .09, all βs > .13, all *t*s > 3.53 and all *p*s < .001. Furthermore, the COVs remained constant with increasing target in all conditions, all R^2s < .02, all βs < .002, all *t*s < 1.35 and all *p*s > .18.

The repeated measures analysis indicated a significant main effect of target, $F(2,80) = 41.89$, *p*GG < .001, $\eta_p^2 = .51$. Linear contrasts showed a significant linear increase in AE score with increasing target (target 12: 3.28, target 18: 5.02, target 27: 7.42), $F(1,40) = 43.83$, $p < .001$, $\eta_p^2 = .52$. The main effect of condition was not significant, $F(2,80) = .13$, $p = .88$, $\eta_p^2 = .003$. The interaction between target and condition did reach significance, $F(4,160) = 4.28$, *p*GG = .02, $\eta_p^2 = .10$. Post-hoc pairwise *t*-tests were performed between the different conditions separately for each target. These indicated significant differences in precision between the multiple sensory control condition on the one hand and the single sensory control,

TABLE 1
Statistics of the regression analyses on the mean estimates, *SDs* of these estimates and COVs to verify scalar variability

	Single sensory control		
	Mean estimates	*SDs*	*COV*
R^2	.41	.10	.01
β	.88	.13	−.001
t	9.15	3.76	−1.10
p	<.001	<.001	.27
	Multiple sensory control		
	Mean estimates	*SDs*	*COV*
R^2	.44	.09	.02
β	.85	.14	−.002
t	9.67	3.53	−1.35
p	<.001	.001	.18
	Congruent		
	Mean estimates	*SDs*	*COV*
R^2	.54	.24	.00047
β	1.04	.19	−.00023
t	11.98	6.16	−.24
p	<.001	<.001	.81

$t(40) = 4.14$, $p < .001$, and the congruent condition, $t(40) = 2.51$, $p = .02$, on the other hand, but only for target 12. This suggests that participants were less precise to estimate target 12 in the multiple sensory control condition (Figure 4).

The correlation analyses indicated significant correlations between the AE scores in all three conditions (single sensory control and multiple sensory control: $r = .87$, $p < .001$, 95% confidence interval (CI) [.768, .928]; single sensory control and congruent: $r = .81$, $p < .001$, 95% CI [.670, .894]: multiple sensory control and congruent: $r = .73$, $p < .001$, 95% CI [.545, .847]).

Discussion

The goal of Experiment 1 was to explore the effect of different methods used to construct arrays of dots on estimates of ANS acuity in numerosity estimation (i.e. the AE score or imprecision of the estimates). In a first condition (single sensory control), the stimuli were created according to the method of Dehaene et al. (2005), whereas in a second condition (multiple sensory control), we used the method of Gebuis and Reynvoet (2011a).

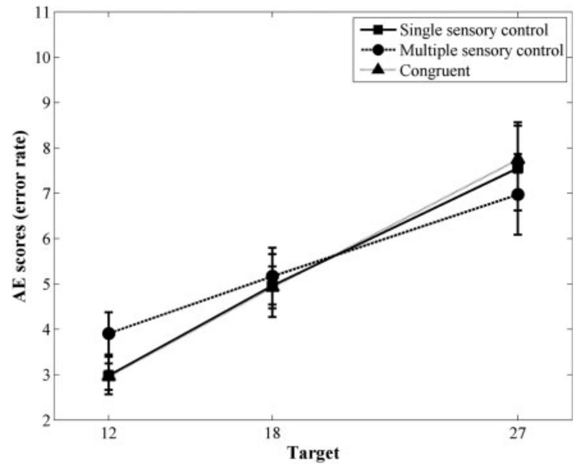

Figure 4. Precision results for the different conditions (single sensory control, multiple sensory control and congruent) as indicated by the AE scores in Experiment 1. Participants performed at a similar level in all three conditions.

In a third condition (congruent), all visual cues co-varied linearly with numerosity.

The results showed the expected signature effect of the ANS in all conditions: the mean estimates and the *SDs* of the participants' estimates increased with increasing target number. This scalar variability is a consequence of the Weber-Fechner law (Fechner, 1860) and was further evidenced by constant COVs with increasing target in all three conditions. These results are in accordance with the findings of previous studies (e.g. Crollen et al., 2011; Mejias, Mussolin, Rousselle, Grégoire, & Noël, 2012; Mejias & Schiltz, 2013).

More crucially however were the effects of the different methods to construct dot arrays, which differ in the manipulation or control of visual cues. We expected that if stable estimates of ANS acuity can be consistently derived within individuals when assessed by means of numerosity estimation, there would be no influence of the specific method that was used to create the dot arrays. More concrete, we especially expected there to be no significant differences in the precision of the estimates (i.e. the AE scores). In fact, participants performed very similarly in all three conditions. Only for the smallest target numerosity (i.e. target 12), the method of creating dot arrays seemed to have a minor effect: participants were slightly less precise for target 12 in the multiple sensory control condition than in the other two conditions. Furthermore, not only were there no substantial differences in AE scores between the three conditions, performances in these three conditions were additionally significantly related. This suggests that all three conditions were supported by the exact same underlying mechanism, leading to highly consistent estimates of ANS acuity across conditions.

In sum, the results of Experiment 1 showed rather small effects of the method employed to construct arrays of dots in an estimation task. This suggests that a stable and consistent estimate of ANS acuity can be obtained within participants by means of this task, even when the stimuli are constructed differently (i.e. the task has sufficient construct validity). Furthermore, this warrants the conclusion that results from estimation studies which used distinct methods to construct the stimuli can be compared and combined to describe the developmental trajectory of ANS acuity. However, the fact that participants' performance was not strongly influenced by the method that was used to construct dot arrays does not necessarily exclude the possibility that these distinct methods may have an effect on performance in numerosity comparison (e.g. Halberda et al., 2008; Moyer & Landauer, 1967; Sasanguie et al., 2012). We hypothesised that estimates of ANS acuity may not be stable when measured with a comparison task in which the non-symbolic stimuli are simultaneously presented on the screen: presenting participants with two stimuli and requiring them to indicate the larger numerosity may encourage them to rely on visual cues because this can be done fairly easy within the design of a simultaneous comparison task. In the estimation task however, a response in symbolic format is required which may probe participants to rely less on these visual cues. It is therefore probable that participants in the comparison task are more affected by the visual cues of the stimuli than is the case in numerosity estimation. This would be evidenced by differences in performance between the three conditions. Additionally, we predicted that correlations between performances in the different conditions of a comparison task would be absent, suggesting that they may be partly directed by distinct underlying mechanisms and that estimates of ANS acuity obtained with the comparison task are not consistent within participants. This possibility was explored in Experiment 2.

EXPERIMENT 2: COMPARISON TASK

Method

Participants. Twenty-six participants took part in Experiment 2 (mean age = 18.62 years, $SD = 1.72$; 21 females). They received a course credit for their participation. The experiment was approved by the Ethical Committee of the Faculty of Psychology and Educational Sciences of the University of Leuven. All participants gave written informed consent.

Stimuli and procedure. The presentation of the stimuli on a 17-inch colour screen and recording of the data was done by means of E-prime 1.1 (Psychology Software Tools, http://www.pstnet.com/). Similar as in Experiment 1, stimuli were white dots on a black background. Participants were instructed to indicate the larger numerosity of two simultaneously presented dot arrays (i.e. a comparison task). Again, there were three conditions: (1) a single sensory control condition with stimuli developed according to the method of Dehaene et al. (2005), (2) a multiple sensory control condition with stimuli constructed with the script of Gebuis and Reynvoet (2011a) and (3) a congruent condition in which the visual cues linearly co-varied with numerosity. Post-hoc analyses in the multiple sensory control condition ensured the absence of a relationship between numerical distance and the difference in visual properties (all R^2 values < .05). The numerosities ranged between 12 and 27 dots. We manipulated the ratio between the two numerosities, resulting in three ratios: ratio 1.5, ratio 1.4 and ratio 1.2. One numerosity always equalled 18, whereas the second numerosity was smaller than 18 in half of the trials (12, 13 and 15) and larger than 18 in the other half of the trials (22, 25 and 27).

On each trial, participants were presented with a fixation cross for 500 ms. Next, two numerosities were presented simultaneously in an invisible grid for 1000 ms and afterwards a blank screen appeared. Participants could respond either during the presentation of the stimuli or during the blank. They were instructed to press the left key when they judged the numerosity on the left side of the screen to be more numerous and the right key when they judged the numerosity on the right side as being more numerous. Each number pair was repeated 16 times, resulting in 32 trials per ratio (2 number pairs per ratio × 16) and 96 trials in total (3 ratios × 32 trials). There was one break in the middle of each condition. The three conditions were administered to all participants following a Latin square distribution.

Data analyses. Mean accuracies were calculated per ratio per participant (3 ratios: 1.2, 1.4 and 1.5).

Additionally, we computed participants' individual Weber fractions (w) as a second measure for ANS acuity in the three visual conditions. This was done based on the specific number pairs (i.e. two number pairs per ratio), resulting in twice the three manipulated ratios, similarly as in Szücs et al. (2013). These symmetrical data are depicted in Figure 6A. While assuming a linear mental number line, model decisions were fitted by means of the following function as described by Piazza et al. (2010) and Halberda et al. (2008):

$$\text{Proportion judged larger } (n1, n2) = \frac{1}{2}\,\text{erfc}\left(\frac{n2 - n1}{\sqrt{2}w\sqrt{n1^2 + n2^2}}\right),$$

where $n1$ is a numerosity that needs to be discriminated from the reference numerosity, $n2$, and erfc is the complementary error function (i.e. a mathematical function). The decision curves were formulated for all Weber fraction values between .01 and 10 in steps of .01. A least squares algorithm identified the model decision curves which best fitted the data of each participant (Szücs et al., 2013; see Figure 6B for the average decision curves per condition). Subsequently, the appropriate individual Weber fraction could be derived separately in each condition. The absence of an effect of the order in which participants were presented with the three conditions was verified for the accuracies, $F(2,23) = .02$, $p = .98$, $\eta_p^2 = .002$. Therefore, presentation order was not taken into account in the following analyses.

The distribution of Weber fractions was found to be non-normal (assessed by means of Kolgomorov–Smirnov statistics for Weber fraction data from all conditions: single sensory control condition: $D = .52$; $p < .001$; multiple sensory control: $D = .55$; $p < .001$; congruent: $D = .50$; $p < .001$). Hence, the data were analysed by non-parametric bootstrap CI estimation instead of regular parametric analyses. In addition, 95% bootstrap CIs were also estimated for the differences in Weber fraction between conditions. Finally, correlation analyses were performed between the three conditions for both the average accuracies and the Weber fractions.

Results

The repeated measures analysis with accuracy as dependent variable indicated a main effect of ratio, $F(2,50) = 96.16$, $p < .001$, $\eta_p^2 = .79$. A linear contrast showed that participants became linearly more accurate when the ratio increased from 1.2 to 1.5, $F(1,25) = 165.55$, $p < .001$, $\eta_p^2 = .87$ (respectively, 84%, 92% and 94%). There was also a main effect of condition, $F(2,50) = 104.27$, $p\text{GG} < .001$, $\eta_p^2 = .81$. Subsequent pairwise t-tests between the average accuracies of the three conditions indicated significant differences in average accuracy between all three conditions (single versus multiple sensory control: $t(25) = 9.96$, $p < .001$; single sensory control versus congruent: $t(25) = -3.90$, $p = .001$; multiple sensory control versus congruent: $t(25) = -10.90$, $p < .001$). Participants performed best in the congruent condition (97%), followed by performance in the single sensory control condition (95%). The worst performance was found in the multiple sensory control condition (79%). Moreover, the factors condition and ratio were included in a significant interaction, $F(4,100) = 3.25$, $p\text{GG} = .03$, $\eta_p^2 = .12$. Separate linear contrast analyses for each condition indicated the presence of a linear increase in accuracy with increasing ratio in all three conditions (single sensory control: $F(1,25) = 70.60$, $p < .001$, 89%, 97%, 98%; multiple sensory control: $F(1,25) = 72.11$, $p < .001$, 71%, 81%, 85%; congruent: $F(1,25) = 42.15$, $p < .001$, 92%, 98%, 99%). Participants thus exhibited a significant ratio effect in accuracy in all conditions. The ratio effect in each condition was further explored by calculating the difference scores between the most difficult ratio (1.2) and the easiest one (1.5) to obtain a measure of the size of the ratio effect (e.g. Defever, Sasanguie, Gebuis, & Reynvoet, 2011; Holloway & Ansari, 2009). Pairwise t-tests on these difference scores indicated a larger difference score in the multiple sensory control condition (13.61%) than in both the single sensory control condition (9.55%), $t(25) = 2.27$, $p =.03$, and the congruent condition (7.62%), $t(25) = 2.59$, $p = .02$. There was no significant difference between the size of the ratio effect of the latter two conditions, $t(25) = 1.62$, $p = .12$ (Figure 5).

In addition, Weber fractions were evaluated by robust, distribution independent bootstrap CI estimations of the Weber fractions themselves (Figure 6C) and bootstrap CI estimations of differences between the three conditions (Figure 6D). The bootstrap CIs were computed from 10,000 random permutations with replacement (Chihara & Hesterberg, 2011). Differences between conditions can be considered significant if the CIs of differences do not include zero (zero meaning no difference between two conditions).

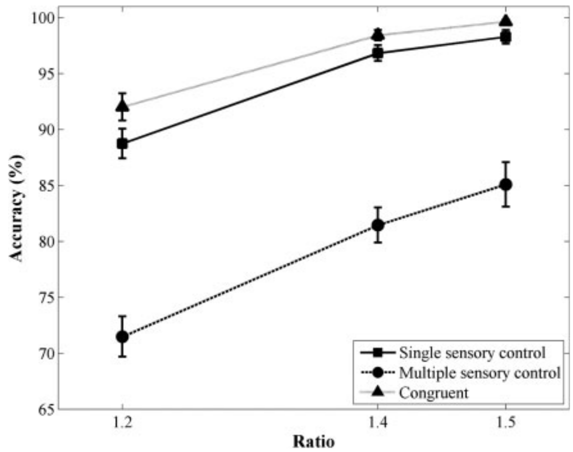

Figure 5. Accuracy results for the different conditions (single sensory control, multiple sensory control and congruent) in Experiment 2. The signature ratio effect was present in all three conditions, as well as differences in average accuracy between the three conditions. Performance was best in the congruent condition, followed by the single sensory control and the multiple sensory control condition.

The non-parametric bootstrap CI estimation indicated that Weber fractions were the largest in the multiple sensory control condition, M = .28, 95% bootstrap CI [.240, .315], followed by the single sensory control condition, M = .11, 95% bootstrap CI [.102, .126]. The smallest Weber fractions were found in the congruent condition, M = .09, 95% bootstrap CI [.070, .010] (Figure 6C). In addition, the 95% bootstrap CIs for the differences of Weber fractions between conditions showed that all Weber fractions differed between all conditions (multiple versus single sensory control: .16, 95% CI [.129, .201]; multiple sensory control versus congruent: .19, 95% CI [.154, .235]; single sensory control versus congruent: .03, 95% CI [.015, .040] (Figure 6D). Considering the strong correlation between accuracies and Weber fractions (single sensory control: $r = -.96$, $p < .001$; multiple sensory control: $r = -.98$, $p < .001$; congruent: $r = -.90$, $p < .001$), these similar results for both are not surprising.

The correlation analyses between average accuracies pointed out that only the correlation between individual accuracies in the single sensory control condition (95%) and the congruent condition (97%) proved to be significant, r = .63, p = .001, 95% CI [.321, .817]. The other two correlations were not significant, indicating the absence of a relationship between average accuracy in the multiple sensory control condition and average accuracy in the single sensory control, r = 23, p = .25, 95% CI [−.172, .566], or the congruent condition, r = .07, p = .73, 95% CI [−.326, .445]. The results of similar correlations between the Weber fractions of the different conditions were in correspondence with the accuracy results: the Weber fractions of the single sensory control condition (.11) and these of the congruent condition (.09) were significantly correlated, r = .57, p = .002, 95% CI [.235, .784], whereas the other two correlations did not reach significance (multiple sensory control versus single sensory control: r = .31, p = .13, 95% CI [−.087, .622]; multiple sensory control versus congruent: r = .05, p = .82, 95% CI [.344, .429]).

GENERAL DISCUSSION

A growing number of studies demonstrating that methodological differences between studies might lead to differences in ANS acuity measures suggest that there is a need to further investigate these methodological issues (e.g. Gebuis & van der Smagt, 2011; Inglis & Gilmore, 2013; Price et al., 2012; Smets et al., 2013; Szücs et al., 2013). This is especially relevant considering the common implicit assumption that estimates of ANS acuity obtained in studies with differing methodologies are comparable and can be combined to describe the developmental trajectory of ANS acuity (e.g. Halberda & Feigenson, 2008; Piazza et al., 2004) as a potential means to detect future mathematical difficulties. In the present study, we focused on the assessment of one essential methodological aspect differing between studies investigating ANS acuity, namely the diverse methods that are used to construct dot arrays in order to control for the visual cues that accompany numerosity. In both numerosity estimation (Experiment 1) and comparison (Experiment 2), we contrasted two methods commonly used in the field of numerical cognition and added a third—more ecologically valid—condition, leading to three conditions: (1) a single sensory control condition in which stimuli were created with the script of Dehaene et al. (2005) (i.e. only one visual cue is kept constant in a single trial), (2) a multiple sensory control condition in which stimuli were created with the script of Gebuis and Reynvoet (2011a) (i.e. manipulation of multiple visual cues) and (3) a congruent condition.

The results of Experiment 1 revealed that in an estimation task, the different methods to construct dot arrays and control their visual cues did not lead

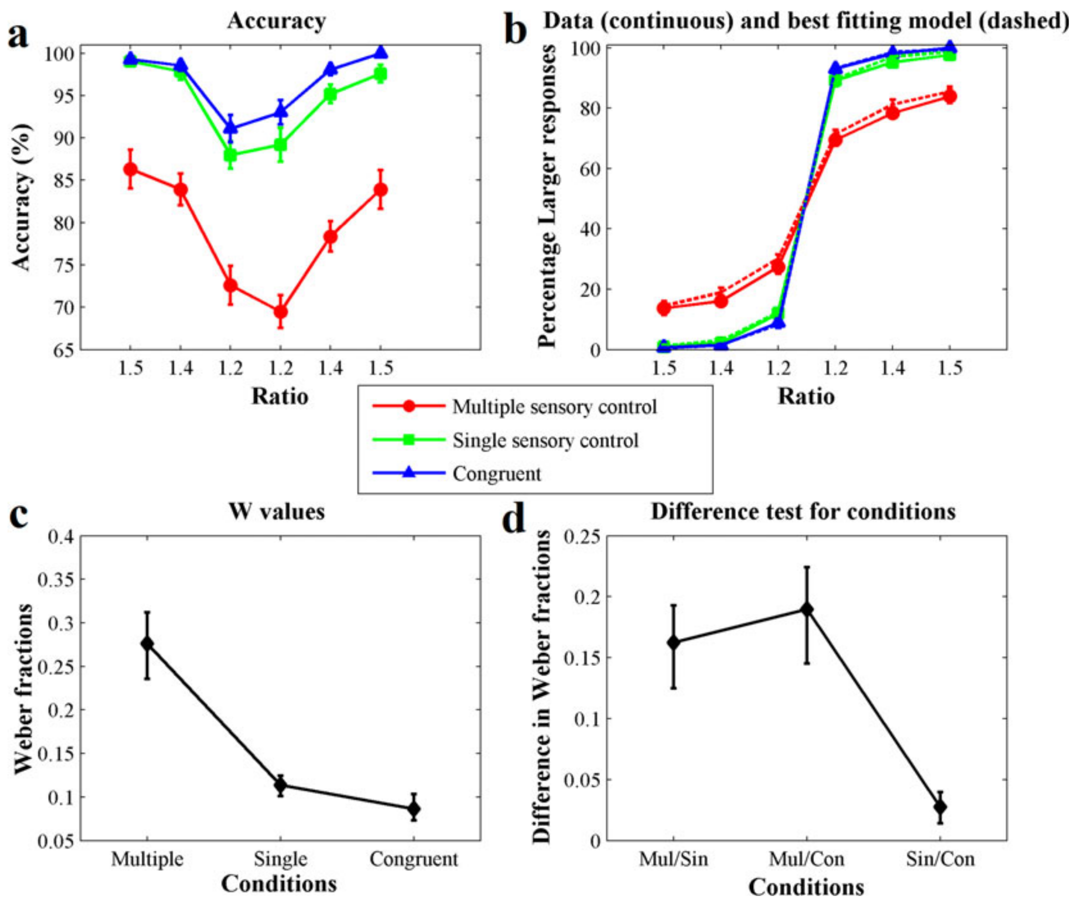

Figure 6. Accuracy and Weber fraction results. The left upper image (a) represents mean accuracy rates with standard errors for the multiple sensory control condition (red line), the single sensory control condition (green line) and the congruent condition (blue line). The right upper image (b) shows the mean empirical decision curves (continuous lines) and the best fitting model decision curves (dashed lines) for the three conditions (standard errors for this curve are the same as for accuracy curves as these curves demonstrate the same data from different points of view). The lower left image (c) exhibits the average Weber fractions with 95% bootstrap CIs of each condition, and the lower right image (d) shows the results of difference tests between each pair of the three conditions with 95% bootstrap CIs. [To view this figure in colour, please visit the online version of this Journal.]

to obvious differences in estimates of ANS acuity (i.e. AE scores or imprecision of participants' estimates): participants' estimation performance was similar in all three conditions and furthermore significantly correlated. This suggests that an estimate of ANS acuity derived from numerosity estimation is consistent within participants and that results from studies which applied a different method to generate the stimuli are comparable.

On the contrary however, the results of Experiment 2 indicated that the different methods to construct dot arrays did cause significant differences in average accuracy and Weber fractions in the comparison task (Experiment 2). The best performance was obtained in the congruent condition, followed by performance in the single sensory control condition. In the multiple sensory control condition participants performed significantly worse than in both above conditions. Furthermore, the correlations between either accuracies or Weber fractions in the multiple sensory control condition on the one hand and the other two conditions on the other hand were not significant and decidedly lower than similar correlations computed in Experiment 1. In addition, these non-significant correlations also did not satisfy the minimum level of acceptable alternate forms or parallel reliability (i.e. .7). This provides evidence that an estimate of ANS acuity derived by means of the commonly used comparison task is neither stable nor consistent within participants, considering that differential methods to construct the dot arrays lead to unrelated performances. Furthermore, the single sensory control condition bears more resemblance in performance to the congruent condition than it does to the multiple sensory control condition, raising doubt about the control of visual cues by means of the first method.

From a theoretical point of view, the results of Experiment 2 may be problematic for the widely accepted global definition of a robust ANS (Feigenson et al., 2004), representing numerosities *independently from visual cues*. Estimates of the acuity of the ANS are therefore presumed to be stable and independent of variations in how the stimuli are constructed, especially within the framework of one task. This would make it possible to compare results from different studies, in particular if Weber fractions are used as an objective measure of ANS acuity in numerosity comparison (Gilmore, Attridge, De Smedt, & Inglis, 2014). However, our results show that at least in a comparison task, participants' performance is strongly influenced by the type of visual cue control at hand, ultimately leading to unrelated accuracies and even Weber fractions.

A potential explanation for the effects of different methods to create the dot arrays, leading to variable estimates of ANS acuity in Experiment 2, could be that participants, when possible, attempt to perform the task based on the visual cues that accompany numerosity. When stimuli are simultaneously presented and participants are instructed to indicate the larger numerosity, relying on visual cues may come naturally because a visual cue comparison can be done rather simply. As a consequence of this (additional) reliance on visual cues, performance will be increased in the congruent and the single sensory control condition, because of the (partly) congruent nature of the visual cues. This is in line with the results of Gebuis and Reynvoet (2011a) and Szücs et al. (2013). Both studies found significant differences in performance on congruent (visual cues congruent with numerosity) and incongruent (visual cues incongruent with numerosity) trials. Additionally, Gebuis and Reynvoet (2011a) have shown that these congruency effects increased when more visual cues were congruent with numerosity. Our results imply that numerosity judgements in a comparison task are facilitated when the visual cues completely (in the congruent condition) or partly (in the single sensory control condition) co-vary with numerosity. This suggestion that visual cues interfere is puzzling when assuming the existence of an ANS that is readily able to extract numerosity independently from visual cues: if visual cues do not influence estimates of numerosity, (small) differences in the methods used to construct dot arrays and control their visual cues should not lead to unrelated performances. Alternatively, it may be more valid to speculate that these particular visual cues play a role from the initial stages of numerosity processing up until the final stages instead of assuming an ANS that is able to extract abstract numerosity independently of these cues. One possibility is that participants integrate visual cue information in order to arrive at a numerosity judgement (Gebuis & Reynvoet, 2011b). This implies that a representation of numerosity may not exist or is merely derived from the integration of several visual cues (Tokita & Ishiguchi, 2013). Another possibility is that numerosity and visual cue information are processed in parallel and combined at a later stage where interference between both might take place (Fuhs & McNeil, 2013; Gilmore et al., 2013). Leibovich and Henik (2013) recently suggested that the discrimination of visual cues and not numerosity is an innate feature. According to their model, children will arrive at an abstract representation of numerosity by gradually learning and internalising the association between visual cues and numerosity throughout development. However, as we observed that this "adult" abstract representation may still be influenced by visual cues, it may be viable to conclude that in addition to relying on the learned abstract representation of numerosity, people continue to rely on visual cues whenever possible. Especially in a comparison task with a simultaneous presentation of numerosities (as opposed to, for instance, a sequential presentation), participants may be significantly probed to depend on visual cues in addition to relying on their abstract representation of numerosity. This reliance on visual cues seems to be less the case in the estimation task, potentially because of the required symbolic numerical output in this task, thus leading to more stable estimates of ANS acuity in this task.

In sum, we can conclude that in numerosity comparison as opposed to estimation, participants' responses are influenced by how the dot arrays are created and as a consequence, how visual cues are manipulated. ANS acuity is not stable and consistent within participants when obtained by means of a comparison task, considering the significantly different and unrelated performances in the three conditions of this task.

Consequently, performances obtained by means of the comparison task are not comparable across different studies using a different method to construct the dot arrays. A potential source of the inconsistency in estimates of ANS acuity may be the differential contribution of visual cues on participants' performance in the different conditions of the comparison task, despite the use of visual cue controls. This result is problematic for the widely accepted view of the ANS as a robust system capable of extracting pure numerosity independently from visual cues: if differential visual cue controls exert an influence on numerosity comparison, discriminating numerosities in this task does not occur completely independently from visual cues. The exact mechanism concerning how exactly visual cues influence numerosity comparison performance cannot be discerned on the basis of the results of the present study. Considering, however, the fact that the comparison task is the most frequently used task in the literature, future research should certainly pay more attention to the construction of dot arrays and their corresponding visual cues, for instance by taking into account the effects of the used method to interpret the results.

REFERENCES

Agrillo, C., Piffer, L., & Bisazza, A. (2011). Number versus continuous quantity in numerosity judgments by fish. *Cognition*, *119*, 281–287. doi:10.1016/j.cognition.2010.10.022

Barth, H., Kanwisher, N., & Spelke, E. (2003). The construction of large number representations in adults. *Cognition*, *86*, 201–221. doi:10.1016/S0010-0277(02)00178-6

Brankaer, C., Ghesquière, P., & De Smedt, B. (2014). Children's mapping between non-symbolic and symbolic numerical magnitudes and its association with timed and untimed tests of mathematics achievement. *PLoS ONE*, *9*(4), e93565. doi:10.1371/journal.pone.0093565.t003

Brannon, E. M. (2006). The representation of numerical magnitude. *Current Opinion in Neurobiology*, *16*, 222–229. doi:10.1016/j.conb.2006.03.002

Castronovo, J., & Göbel, S. M. (2012). Impact of high mathematics education on the number sense. *PLoS ONE*, *7*(4), e33832. doi:10.1371/journal.pone.0033832.t002

Chihara, L., & Hesterberg, T. (2011). *Mathematical statistics with resampling and R*. New York: John Wiley & Sons.

Cordes, S., Gelman, R., Gallistel, C. R., & Whalen, J. (2001). Variability signatures distinguish verbal from nonverbal counting for both large and small numbers. *Psychonomic Bulletin & Review*, *8*, 698–707. doi:10.3758/BF03196206

Crollen, V., Castronovo, J., & Seron, X. (2011). Under- and over-estimation. *Experimental Psychology (formerly Zeitschrift Für Experimentelle Psychologie)*, *58*(1), 39–49.

Defever, E., Reynvoet, B., & Gebuis, T. (2013). Task- and age-dependent effects of visual stimulus properties on children's explicit numerosity judgments. *Journal of Experimental Child Psychology*, *116*, 216–233. doi:10.1016/j.jecp.2013.04.006

Defever, E., Sasanguie, D., Gebuis, T., & Reynvoet, B. (2011). Children's representation of symbolic and nonsymbolic magnitude examined with the priming paradigm. *Journal of Experimental Child Psychology*, *109*, 174–186. doi:10.1016/j.jecp.2011.01.002

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

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

Fechner, G. G. (1860). Elemente der Psychophysik [Elements of psychophysics]. Leipzig: Breitkopf Und Hartel. *Report in James (1890)*, *2*, 50.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences*, *8*, 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*(1), 136–148. doi:10.1111/desc.12013

Gebuis, T., & Reynvoet, B. (2011a). Generating nonsymbolic number stimuli. *Behavior Research Methods*, *43*, 981–986. doi:10.3758/s13428-011-0097-5

Gebuis, T., & Reynvoet, B. (2011b). The interplay between nonsymbolic number and its continuous visual properties. *Journal of Experimental Psychology: General*, *141*, 642–648. doi:10.1037/a0026218

Gebuis, T., & Reynvoet, B. (2012). The role of visual information in numerosity estimation. *PLoS ONE*, *7*(5), e37426. doi:10.1371/journal.pone.0037426.g003

Gebuis, T., & Reynvoet, B. (2013). The neural mechanisms underlying passive and active processing of numerosity. *NeuroImage*, *70*, 301–307. doi:10.1016/j.neuroimage.2012.12.048

Gebuis, T., & van der Smagt, M. J. (2011). False approximations of the approximate number system? *PLoS ONE*, *6*(10), e25405. doi:10.1371/journal.pone.0025405.g003

Gilmore, C., Attridge, N., Clayton, S., Cragg, L., Johnson, S., Marlow, N., … Inglis, M. (2013). Individual differences in inhibitory control, not nonverbal number acuity, correlate with mathematics achievement. *PLoS ONE*, *8*(6), e67374. doi:10.1371/journal.pone.0067374.g002

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

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

Halberda, J., & Feigenson, L. (2008). Developmental change in the acuity of the "Number Sense": The approximate number system in 3-, 4-, 5-, and 6-year-olds and adults. *Developmental Psychology*, *44*, 1457.

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*, 11116–11120. doi:10.1073/pnas.1200196109

Halberda, J., Mazzocco, M. M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature*, *455*, 665–668. doi:10.1038/nature07246

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

Inglis, M., & Gilmore, C. (2013). Sampling from the mental number line: How are approximate number system representations formed? *Cognition*, *129*(1), 63–69. doi:10.1016/j.cognition.2013.06.003

Inglis, M., & Gilmore, C. (2014). Indexing the approximate number system. *Acta Psychologica*, *145*, 147–155. doi:10.1016/j.actpsy.2013.11.009

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

Katz, C., & Knops, A. (2014). Operational momentum in multiplication and division? *PLoS ONE*, *9*(8), e104777. doi:10.1371/journal.pone.0104777.t003

Leibovich, T., & Henik, A. (2013). Magnitude processing in non-symbolic stimuli. *Frontiers in Cognition*, *4*, 375.

Libertus, M. E., & Brannon, E. M. (2010). Stable individual differences in number discrimination in infancy. *Developmental Science*, *13*, 900–906. doi:10.1111/j.1467-7687.2009.00948.x

Linsen, S., Verschaffel, L., Reynvoet, B., & De Smedt, B. (2014). The association between children's numerical magnitude processing and mental multi-digit subtraction. *Acta Psychologica*, *145*, 75–83. doi:10.1016/j.actpsy.2013.10.008

Lyons, I. M., Price, G. R., Vaessen, A., Blomert, L., & Ansari, D. (2014). Numerical predictors of arithmetic success in grades 1–6. *Developmental Science*, *17*, 714–726. doi:10.1111/desc.12152

Meck, W. H., & Church, R. M. (1983). A mode control model of counting and timing processes. *Journal of Experimental Psychology: Animal Behavior Processes*, *9*, 320–334. doi:10.1037/0097-7403.9.3.320

Mejias, S., Grégoire, J., & Noël, M.-P. (2012). Numerical estimation in adults with and without developmental dyscalculia. *Learning and Individual Differences*, *22*(1), 164–170. doi:10.1016/j.lindif.2011.09.013

Mejias, S., Mussolin, C., Rousselle, L., Grégoire, J., & Noël, M.-P. (2012). Numerical and nonnumerical estimation in children with and without mathematical learning disabilities. *Child Neuropsychology*, *18*, 550–575. doi:10.1080/09297049.2011.625355

Mejias, S., & Schiltz, C. (2013). Estimation abilities of large numerosities in kindergartners. *Frontiers in Developmental Psychology*, *4*, 518.

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

Pahl, M., Si, A., & Zhang, S. (2013). Numerical cognition in bees and other insects. *Frontiers in Psychology*, *4*, 162. doi:10.3389/fpsyg.2013.00162

Piazza, M., Facoetti, A., Trussardi, A. N., Berteletti, I., Conte, S., Lucangeli, D., … Zorzi, M. (2010). Developmental trajectory of number acuity reveals a severe impairment in developmental dyscalculia. *Cognition*, *116*(1), 33–41. doi:10.1016/j.cognition.2010.03.012

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

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

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

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

Restle, F. (1970). Speed of adding and comparing numbers. *Journal of Experimental Psychology*, *83*, 274–278. doi:10.1037/h0028573

Revkin, S. K., Piazza, M., Izard, V., Cohen, L., & Dehaene, S. (2008). Does subitizing reflect numerical estimation? *Psychological Science*, *19*, 607–614. doi:10.1111/j.1467-9280.2008.02130.x

Sasanguie, D., Defever, E., Van den Bussche, E., & Reynvoet, B. (2011). The reliability of and the relation between non-symbolic numerical distance effects in comparison, same-different judgments and priming. *Acta Psychologica*, *136*(1), 73–80. doi:10.1016/j.actpsy.2010.10.004

Sasanguie, D., Göbel, S. M., Moll, K., Smets, K., & Reynvoet, B. (2012). Approximate number sense, symbolic number processing, or number–space mappings: What underlies mathematics achievement? *Journal of Experimental Child Psychology 114*, 418–431. doi:10.1016/j.jecp.2012.10.012

Schmidt, R. A., & Lee, T. (2011). *Motor control and learning: A behavioral emphasis*. Champaign, IL: Human Kinetics.

Smets, K., Gebuis, T., & Reynvoet, B. (2013). Comparing the neural distance effect derived from the non–symbolic comparison and the same–different task. *Frontiers in Human Neuroscience*, *7*, 28. doi:10.3389/fnhum.2013.00028

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

Szücs, 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.

Tokita, M., & Ishiguchi, A. (2013). Effects of perceptual variables on numerosity comparison in 5–6-year-olds and adults. *Frontiers in Psychology*, *4*, 431. doi:10.3389/fpsyg.2013.00431