Acta Psychologica 210 (2020) 103163

Contents lists available at ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/locate/actpsy

The impact of set size on cumulative area judgments

Sophie Savelkouls*, Sara Cordes

Department of Psychology, Boston College, United States of America

ARTICLE INFO

Keywords:
Continuous quantity
Discrete quantity
Numerical cognition
Interference

ABSTRACT

The ability to track number has long been considered more difficult than tracking continuous quantities. Evidence for this claim comes from work revealing that continuous properties (specifically cumulative area) influence numerical judgments, such that adults perform worse on numerical tasks when cumulative area is incongruent with number. If true, then continuous extent tracking abilities should be unimpeded by number. The aim of the present study was to determine the precision with which adults track cumulative area and to uncover the process by which they do so. Across two experiments, we presented adults with arrays of dots and asked them to judge the relative cumulative area of the displays. Participants performed worse and were slower on incongruent trials, in which the more numerous array had the smaller cumulative area. These findings suggest that number interferes with continuous quantity judgments, and that number is at least *as salient* as continuous variables, undermining claims in the literature that continuous properties are easier to represent, and more salient to adults. Our primary research question, however, pertained to how cumulative area representations were impacted by set size. Results revealed that the area of a single item was tracked much faster and with greater precision than the area of multiple items. However, for sets with more than one item, results revealed less accurate, yet faster responses, as set size increased, suggesting a speed-accuracy trade-off in judgments of cumulative area. Results are discussed in the context of two distinct theories regarding the process of tracking cumulative area.

1. Introduction

Representing quantity is an important skill for human and non-human animals alike. Whether you are a human deciding just how many apples you will need to make your favorite apple pie, or a mosquitofish deciding where in the ocean you can find the highest density of zooplankton, the ability to represent approximate quantities is important for day-to-day life. However, which quantities we rely upon for these important decisions has been a topic of debate (Gebuis & Reynvoet, 2012a; Leibovich et al., 2017; Savelkouls & Cordes, 2017). Human and non-human animals can represent discrete quantity (i.e. number; Humans: e.g., Halberda & Feigenson, 2008; Non-human animals: e.g., Brannon & Terrace, 1998; Meck & Church, 1983) but they can also track continuous quantities (also referred to as continuous extent[1]) such as area, volume, length or density (Humans: e.g., Brannon et al., 2006; Odic, 2018; Marchant, Simons, & de Fockert, 2013; Non-human animals: e.g., Boysen et al., 2001). Not surprisingly, these discrete and continuous quantities are strongly correlated with one another: e.g., 10 apples are not only more numerous than 5 apples, but their cumulative volume, weight, surface area, and density are also greater. This naturally strong correlation between discrete and continuous variables has led researchers to question the extent to which we track these quantities independently of each other.

A majority of the research investigating humans' quantitative abilities has focused on our ability to represent discrete number. While substantial research has supported the idea that infants, children, and adults are remarkably good at representing number (Halberda & Feigenson, 2008; Odic et al., 2015; Xu & Spelke, 2000), not everyone agrees. Proponents of the “Sense of Magnitude” (SoM) theory take a neo-Piagetian approach to number representation, suggesting that our abilities to track number are fully reliant upon an ability to track continuous quantities (Gebuis & Reynvoet, 2012b; Leibovich et al., 2017). The premise of this argument is that continuous quantities such as element area (EA), cumulative area (total area of all items in an array; CA) or density (the ratio of the number/area of items and the size of the display) are derived from, and dependent upon, the perceptual qualities of the display, and thus are significantly easier to track than number. In contrast, number is considered to be an abstract quantity – it

* Corresponding author at: Department of Psychology, Boston College, 140 Commonwealth Avenue, 300 McGuinn Hall, Chestnut Hill, MA 02467, United States of America.

E-mail address: sophie.savelkouls@gmail.com (S. Savelkouls).

[1] Note: For the purposes of this paper, “continuous quantity” will exclusively refer to visual quantities, not e.g., time.

https://doi.org/10.1016/j.actpsy.2020.103163
Received 4 April 2019; Received in revised form 12 August 2020; Accepted 13 August 2020
Available online 25 August 2020
0001-6918/

can be tracked using many different sensory modalities (vision, sound and even touch), and even compared across modalities (e.g. it is possible to compare the number of voices heard to the number of people seen). As such, the ability to track number has been considered to involve much higher order cognitive processes than tracking continuous quantities, making it unlikely that number is tracked in the presence of other perceptual quantities.

As a direct test of these claims, researchers have investigated whether we can track number independent of continuous properties. Although substantial work reveals that continuous properties can bias numerical judgments (Gebuis & Reynvoet, 2012a; Hurewitz et al., 2006; Leibovich et al., 2015), researchers have successfully developed paradigms that systematically control for continuous properties that typically correlate with number, providing strong evidence that humans are capable of tracking number independent of continuous perceptual variables (Halberda & Feigenson, 2008; Lipton & Spelke, 2003; Lourenco et al., 2012; Odic et al., 2013; Starr et al., 2013; Xu et al., 2005; Xu & Spelke, 2000). However, less work has explored the converse; that is, how well can we track continuous properties independent of number? Are continuous properties easier to track? Notably, as posited by the SoM theory, numerical abilities would only be dependent upon continuous extent tracking if and only if continuous extent representations are more precise and more salient than numerical ones. If our abilities to track continuous properties are less refined than our abilities to track number, then it seems unlikely that humans would primarily rely upon less precise continuous representations when tracking multiple items. Although a myriad of studies have examined number discrimination abilities in the context of competing continuous extent information (Barth, 2008; Gebuis & Reynvoet, 2012a; Hurewitz et al., 2006; Leibovich & Henik, 2014; Salti et al., 2017), little research has directly examined our abilities to discriminate arrays on the basis of continuous quantity in the context of competing numerical information.

The aim of the present study was to examine adult abilities to discriminate cumulative area (CA) across arrays with differing numerical information, in order to shed light on the process by which adults track cumulative area. Although there are many different continuous properties we could have chosen to focus on, the current investigation focused exclusively on cumulative area to build upon several previous studies. These studies have revealed that humans are sensitive to changes in the surface area of items in a display and that cumulative area changes may impact number discrimination performance (e.g., Barth, 2008; Cordes & Brannon, 2008; DeWind & Brannon, 2012; Gebuis & Reynvoet, 2012a; Hurewitz et al., 2006). We had two goals: (1) to determine the precision with which adults track CA across various set sizes and (2) to investigate if and how CA discriminations are influenced by numerical information. That is, we aimed to understand the process by which we track continuous quantities – are CA representations dependent upon number? This latter question is of theoretical importance because it can speak to how our representations of number and continuous quantities may be related.

How well do adults discriminate Cumulative Area?

Most studies that have examined adult CA tracking abilities have used discrimination or numerical Stroop type tasks. In these types of tasks, on some trials continuous properties and number are incongruent with one another (e.g. the array with a greater number of items has a smaller CA) and on other trials, number and continuous properties are congruent (e.g. the array with the greater number of items also has a greater CA). Although explicit instructions are to judge the relative numerosity of the displays (i.e., not to attend to cumulative area), multiple studies have found that adults consistently perform worse on incongruent compared to congruent trials (Barth, 2008; DeWind & Brannon, 2012; Gebuis & Reynvoet, 2012a; Hurewitz et al., 2006). This has been taken as evidence that even when given explicit instructions to pay attention to number, adults *automatically* process continuous properties of the set (even when irrelevant to the task). These findings have provided the basis for claims that continuous perceptual properties are more readily and precisely tracked than number.

However, is it true that adults track CA with relatively greater precision than that of number? If numerical judgments were fully dependent upon continuous properties, then one would expect our ability to track continuous quantities to be more refined than that of number. That is, humans should be at least as good at discriminating arrays based upon continuous quantities as they are at discriminating arrays based upon number. However, very few studies have specifically examined adult abilities to discriminate continuous properties. A handful of studies have compared the precision with which infants, children, and adults track the *area of a single item* (e.g., a circle or amorphous shape) to their numerical tracking abilities, reporting similar, or even more precise area tracking abilities (Anderson & Cuneo, 1978; Brannon et al., 2006; Leibovich & Henik, 2014; Lourenco & Bonny, 2017; Odic et al., 2013). Yet, critically, the only way to address claims (such as those made by SoM theory) that continuous extent is more readily tracked over number is to examine the precision of area representations in the context of *sets* where numerical information is available. Importantly, studies that have included trials in which number and area are incongruent with one another have found performance costs on incongruent trials such that CA judgments are less accurate when number is incongruent with CA (Barth, 2008; Hurewitz et al., 2006).[2] Other studies that have presented participants with sets of items but kept the number of items identical in both sets found that participants performed just as well on the area task compared to a number task (Lourenco et al., 2012). That is, number may be just as salient as CA judgments in the context of sets.[3]

There is some evidence to suggest that continuous quantity discriminations may be less precise than numerical discriminations, at least in human infants. Studies that have investigated 6–7-month-olds' abilities to track the size of individual objects (element area, or EA) when presented in the context of an array of items, or the CA of an array of items, have found that infants needed as much as a 1:4 ratio of change in both EA and CA to detect a change. This 1:4 ratio of change is notably greater than the 1:2 ratio of change necessary to detect changes in number (Brannon et al., 2004; Cordes & Brannon, 2008, 2011). In sum, the infant literature suggests that infants are better at tracking number than continuous dimensions in the context of sets (i.e., more than one item). Which opens up the question – if number interferes with an infant's abilities to track CA, by exactly what process is it that we track CA?

Processes involved in CA representation

So how do we represent CA when presented with an array of items? The previous literature has presented us with two mutually exclusive theoretical possibilities.

One possibility, which we will refer to as the 'Direct Perception' hypothesis, is consistent with assumptions of the Sense of Magnitude

[2] Barth (2008) ran additional models of the data and concluded that this decreased performance in incongruent trials was not due to interference between the two dimensions of number, but instead could be explained by the fact that participants underestimated individual element areas resulting in more difficult discriminations. Importantly, we highlight that these findings are consistent with claims that area judgments are more difficult than numerical ones.

[3] Note Hurewitz et al. (2006) reported CA interfered with numerical judgments more so than number interfered with CA judgments. Importantly, however, the authors did not systematically match the ratio of change across the two dimensions, such that numerical differences across displays may have been significantly smaller (and thus less salient) than CA differences. As such, it is inappropriate to draw conclusions regarding the relative salience of the two quantities from their data.

(SoM) theory. According to the Direct Perception hypothesis we are able to track surface area directly from the perception – that is, we directly perceive exactly how much area is covered just as readily as we notice the color or luminance of the items. Importantly, this direct abstraction does not require extensive cognitive processing, such as summing surface area across individual items, and thus does not require individuating items in the array. As such, the number of items in the display – that is the number of items over which CA is tracked – is irrelevant to CA tracking because it involves a single cognitive process and thus set size should have no impact on the precision of CA acuity, nor the speed with which CA is tracked (i.e., it should take the same amount of time to process the CA of a set of 2 items as it does to process the CA of a set of 15 items). Moreover, it is also conceivable (though not a requirement under this hypothesis) that the precision with which we track the area of a single item may in fact be comparable to the precision with which we track the CA of 20 items because again, the areas in both cases are simply directly perceived.

Support for the Direct Perception hypothesis comes from studies revealing similar infant CA discrimination abilities for small sets (2–3 items) as that of large sets (10–15 items),[4] suggesting that CA acuity is unaffected by set size (Cordes & Brannon, 2008). This Direct Perception Hypothesis underlies many neo-Piagetian claims (e.g. Gebuis & Reynvoet, 2012b; SoM: Leibovich et al., 2017; Mix et al., 2002) positing that continuous extent quantities are directly perceived, at least early in development, making them easier to track than more abstract quantities, like number.

On the other hand, according to the 'Computation' hypothesis (see Barth, 2008), rather than representing CA directly, we may track CA by representing the surface area of individual items within an array (likely through direct perception of the surface area of individual items) and then summing across these representations (e.g. adding representations of individual areas together). Because prior research suggests that mental summation is not a completely noiseless process (Cordes et al., 2007; McCrink et al., 2007), each addition process is expected to contribute noise to the representation. Thus, precision in the representation of CA should decrease as the number of elements in the display increases. Moreover, because each additional item requires additional processing, arrays with a greater number of elements should take more time to process than ones with fewer elements.

Support for the Computation hypothesis is found in prior research revealing infants are significantly better (i.e., more precise) at tracking the area of a single item, compared to tracking the CA of multiple items (Brannon et al., 2004; Brannon et al., 2006; Cordes & Brannon, 2008). Additionally, Barth (2008) compared quantitative models of adult CA judgments and determined that a summation account provided the best explanation for the data.

The current study

Importantly, no studies have directly examined the effect of set size on cumulative area tracking abilities in adults. With supporting evidence for both the Direct Perception and Computation hypotheses, it is still unclear how numerical information may or may not influence our ability to track CA. One key distinction between the two hypotheses is in the role that number plays in CA discriminations. While the Direct Perception hypothesis assumes that number should have no effect on CA acuity or reaction times (RTs), the Computation hypothesis predicts less CA acuity and increasing RTs with increasing set size.

In the current study, we presented adults with pairs of dot arrays and asked them to judge which array had the greater CA. We manipulated 4 variables: the CA Ratio (the ratio between the CAs of two displays – a way of varying the relative difficulty of the comparison to provide a means of assessing CA acuity), Congruency (whether the display with the larger number of dots had the smaller or greater CA), Number Ratio (the ratio between the number of items in the two displays), and Set Size (how many items were in each display). We explored the following issues:

(1) *Numerical Congruency:* Building on prior research (Barth, 2008; Hurewitz et al., 2006), we aimed to provide clarity on how conflicting or consistent numerical information impacts CA tracking abilities. If (according to the SoM theory) CA is relatively more salient than number, then CA acuity should be unaffected by whether or not numerical information is congruent or incongruent with CA magnitude. On the other hand, some prior work suggests that numerical information is automatically processed, even in the context of a CA judgment task (Barth, 2008; Hurewitz et al., 2006), suggesting both quantities may be similarly salient. If so, then congruent numerical information should promote CA discrimination performance and/or incongruent numerical information should hinder CA discrimination performance.

Moreover, in this study, we purposely manipulated the relative salience of the numerical differences across arrays to determine whether larger numerical differences produce greater congruency effects. To investigate this question, we presented three distinct Number Ratios (1 1.33 and 1.5) across trials to explore how numerical differences between the displays may make number more or less salient in the case of CA judgments. Importantly, through the inclusion of 1 Number ratio trials (i.e., in which the number of items in each of the two arrays was the same), we were also able to compare performance on these number neutral trials to those of congruent and incongruent trials to determine whether numerical congruency facilitates, and/or incongruency hinders, CA judgments.

(2) *Set Size:* The question of how cumulative area judgments are influenced by set size has not previously been addressed. Importantly, understanding the influence of set size – in particular, whether the precision and speed of CA judgments decreases with increasing number - is key to distinguishing between the Direct Perception and Computation hypotheses. To address this, we presented participants with pairs of arrays composed of either: two single dots, small sets (4–7 total dots), medium sets (12–15 total dots), or large sets (20–25 total dots). Importantly, given that the question of Set Size is independent of the question of numerical congruency, we focused this analysis exclusively on trials in which the number of items in the two arrays was the same (i.e., number neutral trials).

Across two experiments, we addressed these research questions by asking adult participants to rapidly judge which of two simultaneously presented visual arrays had the larger CA. Importantly, number was irrelevant to our task demands and should have had no influence on performance.

2. Experiment 1

Given that few studies have examined adult CA judgments, Experiment 1 examined the effect of Set Size on adults' CA discrimination performance. Participants completed a discrimination task in which they were asked to choose which of two arrays of dots had a greater CA.

[4] This study relied upon a standard habituation looking-time paradigm, revealing that infants were capable of discriminating a 4-fold, but not a 3-fold, change in CA, across exclusively small and exclusively large sets. Because infant habituation techniques do not lend themselves to fine-grained assessments of discrimination abilities, it is not possible to determine whether discrimination capabilities may have varied somewhat as a function of set-size between the 4-fold and 3-fold changes.

2.1. Methods

2.1.1. Participants

Seventy-eight Boston College students participated in our study in exchange for cash or course credit (66 female, M = 18.89 years, *Range* = 18–26 years). Informed consent was obtained from all participants.

2.1.2. Procedure

Each participant completed the study on a computer with a 22″ monitor. Participants were first presented with an instruction screen that informed them that on each trial, they should choose the display of dots with the "greater amount of blue, therefore the greater cumulative area of blue." Each trial consisted of two side-by-side displays of blue dots and participants made a forced choice judgment about the presented pair of displays by pressing the left or right arrow key on the keyboard. They were first presented with a minimum of four practice trials; only once they had responded correctly on three of the four practice trials did they move onto the test trials (all participants moved to the test trials after the first set of practice trials). The practice trials were designed to be very easy for the participant, with the CA Ratio changing 3-fold across the two displays, and the number of items (set size) varying from 2 to 12 dots (this was identical in range to the test trials). Next, participants received 190 test trials, with a break every 50 trials (3 breaks total). Participants were encouraged to look away from the computer and talk to the experimenter during the break. The order of the trials, as well as which display was presented on the left or right side of the screen was randomized for each participant.

Across trials we manipulated the following variables: CA Ratio (1.15, 1.33, 1.45, 1.6, or 1.9), Number Ratio (1, 1.33 or 1.5), Congruency (Congruent or Incongruent), and Set Size (Small (4–7 total items), Medium (12–15 total items) or Large (20–25 total items) sets) to determine how these factors influenced participants' CA judgments (See Table 1 for breakdown of trials). Participants experienced a total of 180 trials involving arrays of multiple items: 5 CA ratios x 3 Number Ratios x 2 Congruency × 3 Set Sizes x 2 trials. Importantly, however, trials involving the Number Ratio 1 were neutral trials since they were neither congruent nor incongruent. For these neutral trials, we had one comparison type for each Set Size (i.e., Small Sets: 2v2, Medium Sets: 6v6 and Large Sets: 10v10). Lastly, for each CA Ratio, we included two trials that were "single" trials in which each display contained only one dot (10 single item trials: 2 trials × 5 CA ratios). These trials would allow us to compare participants' area discriminations involving a single item to CA discriminations involving multiple items (trials with small, medium and large set sizes). In total, participants completed 190 trials.

Upon completing the test trials, participants were asked two questions regarding their strategies for performing in the task. Given that participants' responses on these two questions were not related to their actual performance on the task, these questions were dropped from further analysis.

Table 1
Experiment 1: Breakdown of the number of stimuli per variable manipulated.

Number Ratio	1, 1.33 or 1.5						Single
Set Size	Small		Medium		Large		Trials
Congruency	Cong.	Incong.	Cong.	Incong.	Cong.	Incong.	

Participants were tested on 5 CA Ratios (1.15, 1.33, 1.45, 1.6, or 1.9). Then for each CA Ratio, trials were broken down as illustrated above for the variables of Number Ratio, Set Size and Congruency. This resulted in a total of 180 multiple items trials (5 CA ratios x 3 Number Ratios x 2 Congruency × 3 Set Sizes x 2 trials = 180 trials). Additionally, participants were tested on 2 single trials per CA Ratio, for a total of 10 single trials. In total, participants therefore were presented with 190 trials (180 multiple items trials +10 single trials).

2.1.3. Stimuli

All stimulus parameters are available on OSF (https://osf.io/ejb9p/?view_only=972b5df13a9548de89cdc39c46e7fbef). Stimuli were created using Adobe Illustrator (See Fig. 1 for example stimuli). For our CA values, we generated a list of 12 random CA values between 20 and 45 cm^2 (this range was deemed reasonable for our display size, ensuring that each individual dot would not become so small that they would be hard to see, or so big that they would not fit within the stimulus background) and used these 12 values for each Number and CA Ratio. These random numbers were then multiplied by the appropriate CA Ratio to determine the CA of the comparison display. Thus, across all CA Ratios tested, the CA values ranged between 23 and 83.6 cm^2.

To ensure that participants would not be able to use the size of individual dots as a cue for discrimination, the dots in each array were heterogeneous in size. The individual dot sizes were randomly chosen to fall within 35% of the average element area (as per Lidz et al., 2011). Moreover, item density was controlled such that the dots were randomly placed within an invisible rectangular background that varied in size dependent upon the number of items. To do this, the ratio between the number of dots in the display and the size of the background was computed for one display and used to determine the size of the background for its pair to keep item density within trials. Across trials, the dot density varied from 0.009–0.03 items/cm^2 (thus, the item background ranged from 226.7 cm^2 to 460 cm^2).

2.1.4. Data processing and analyses

We performed all analyses on both accuracy (percentage correct) and reaction time (RT). For accuracy analyses, participants whose performance fell 3 standard deviations above or below the mean for their overall percentage correct on the task were excluded from any analyses (N = 1). We also calculated Weber fractions for each participant using only the data from the trials with sets greater than 1. Weber fractions (w) are defined as the smallest change between two quantities that can be reliably be detected. We estimated w using a psychophysical model using Gaussian random variables as has been done in previous research (Halberda & Feigenson, 2008; Izard et al., 2008; Moyer & Bayer, 1976). In short, we inputted each participants' accuracy on the four hardest CA Ratios (1.15, 1.3, 1.45 and 1.6) and manipulated a single free parameter w until we found a Weber curve that best fitted the data and that minimized error (Halberda & Feigenson, 2008).

For RT analyses, only trials with correct responses were included. All RTs 3 standard deviations above or below the average RT for that participant were excluded. Similar to accuracy analyses, participants whose performance fell 3 standard deviations above or below the mean for their overall RT on the task were excluded from any analyses (N = 1).

2.2. Results & discussion

On average, participants performed well on our task with 85.76% accuracy (*Range* = 60.53–98.42%). The average Weber fraction across all trials with multiple items was w = 0.23, (*Range* = 0.07–1.2; SD = 0.16, SE = 0.018). We compared this Weber fraction to the Weber fraction previously reported by Odic et al. (2013) with adults on a number discrimination task (w = 0.13, SE = 0.02, SD = 0.057, N = 8). Since the two samples varied so widely in their standard deviations and the number of participants tested, we conducted an independent samples t-test assuming unequal variances using the Welch-Satterthwaite procedure for unequal variances. This revealed that the weber fraction for our CA discrimination task was significantly higher than that reported for numerical discriminations, $t(16.22) = 3.35$, $p < .01$, suggesting that CA acuity in our study was significantly worse than prior reports of numerical acuity (a higher weber fraction indicates lower acuity).

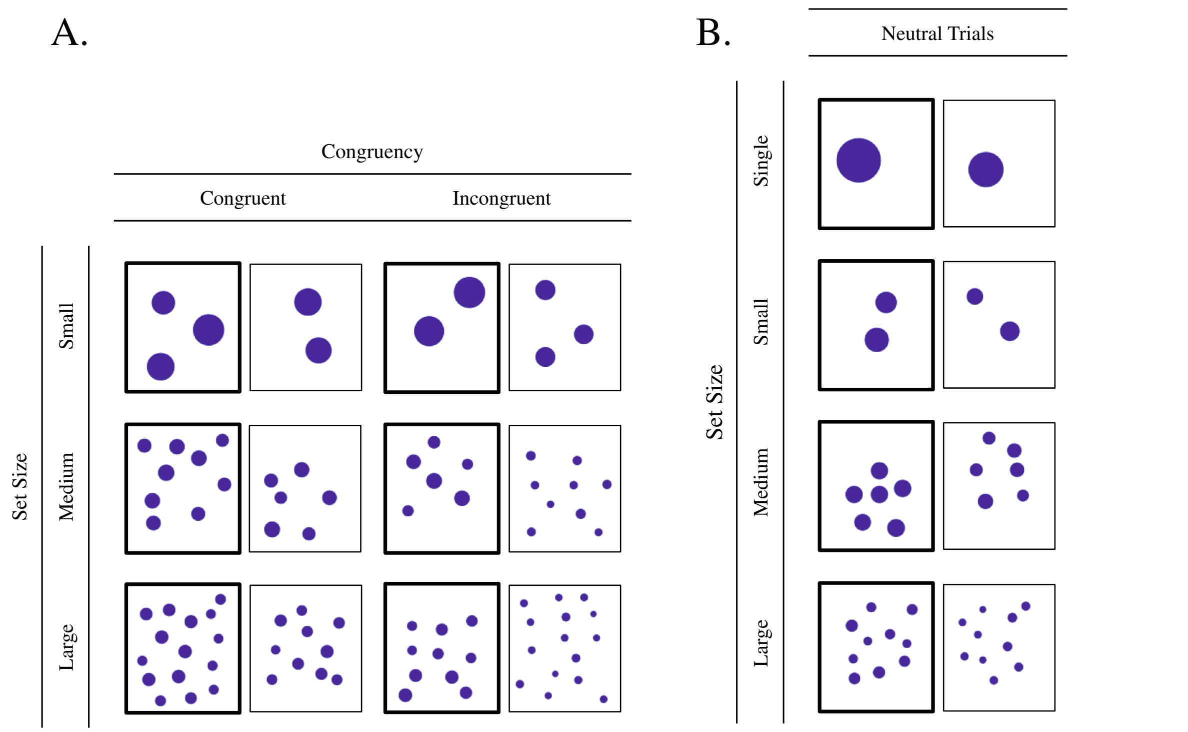

Fig. 1. Example stimuli pairs for the 1.6 CA Ratio.
Fig. 1A. Stimuli are broken down by Set Size (Small, Medium or Large) and Congruency (Congruent or Incongruent). All stimuli are of the 1.33 Number Ratio. For each stimuli pair, the image on the left with the darker border has the larger CA.
Fig. 1B. This figure shows trials involving the Number Ratio 1 (i.e., the number neutral trials). Stimuli are broken down by Set Size, including single item trials.

2.2.1. Does numerical congruency matter for CA judgments?

2.2.1.1. Accuracy. To explore the impact of numerical congruency, we conducted an ANOVA on accuracy data from those test trials involving arrays with more than one item (excluding single item trials) and in which the number of items differed across arrays (i.e., where numerical ratio did not equal one).[5] We ran a 2 (Number Ratio:1.33 or 1.5) x 2 (Congruency: Congruent vs. Incongruent) repeated measures ANOVA on data from those trials. Analyses revealed a main effect of congruency ($F(1,76) = 87.62$, $p < .001$, $\eta_p^2 = 0.54$), such that participants performed significantly better on congruent ($M = 93.55\%$) compared to incongruent trials ($M = 71.30\%$), in line with previous research (Barth, 2008; Hurewitz et al., 2006) and in contradiction of predictions of the SoM Theory. Analyses also revealed a main effect of Number Ratio ($F(1,76) = 20.68$, $p < .001$, $\eta_p^2 = 0.21$) which was qualified by a significant Number Ratio x Congruency interaction ($F(1,76) = 16.03$, $p < .001$, $\eta_p^2 = 0.17$). Although adults performed better on congruent compared to incongruent trials for both numerical ratios (1.33 ratio: $t(76) = 8.26$, $p < .001$, $d = 0.94$; 1.5 ratio: $t(76) = 9.73$, $p < .001$, $d = 1.11$), the impact of congruency was greater when the numerical difference between the arrays was greater ($M_{difference}$ *1.33 ratio* $= 19.61\%$, $M_{difference}$ *1.5 ratio* $= 24.89\%$; $t(76) = 4.00$, $p < .001$, $d = 0.46$; see Fig. 2). Furthermore, although there was no difference in performance between the two Number Ratios for congruent trials ($M_{difference} = 0.52\%$; $t(76) = 0.81$, $p = .41$, $d = 0.09$) there was for incongruent trials ($M_{difference} = 5.80\%$, $t(76) = 4.85$, $p < .001$, $d = 0.55$).

2.2.1.2. Reaction time. We conducted the same 2 (Number Ratio:1.33 or 1.5) x 2 (Congruency: Congruent vs. Incongruent) repeated measures ANOVA but this time looking at reaction times. We again found a main effect of congruency ($F(1,74) = 30.26$, $p < .001$, $\eta_p^2 = 0.29$), such that reaction times were significantly faster on congruent ($M = 946.31$ s) compared to incongruent trials ($M = 1087.57$ s). However, unlike our findings with accuracy, we did not find a significant main effect of Number Ratio ($F(1,74) = 1.68$, $p = .20$, $\eta_p^2 = 0.02$). We did find a significant interaction ($F(1,74) = 12.22$, $p < .001$, $\eta_p^2 = 0.14$). Reaction times were faster on congruent compared to incongruent trials for both numerical ratios (1.33 ratio: $t(74) = 5.95$, $p < .001$, $d = 0.94$; 1.5 ratio: $t(74) = 3.59$, $p < .001$, $d = 1.11$). However, the impact of congruency was greater when the numerical ratio greater ($M_{difference}$ *1.33 ratio* $= 93.87$ s, $M_{difference}$ *1.5 ratio* $= 188.65$ s; $t(74) = 3.50$, $p < .001$, $d = 0.40$). Furthermore, there was a significant difference in performance between the two Number Ratios for congruent trials ($M_{difference} = 63.67$ s; $t(74) = 4.07$, $p < .001$, $d = 0.47$) but not for incongruent trials ($M_{difference} = 31.12$ s; $t(74) = 1.49$, $p < .001$, $d = 0.17$).

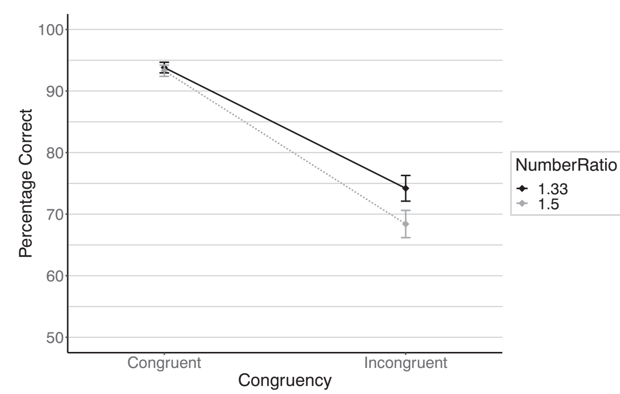

Fig. 2. Experiment 1: percent correct as a function of number ratio (1.33 and 1.5) and congruency (congruent and incongruent). Error bars represent standard error.

[5] Trials involving a Number Ratio of 1 were excluded because there was no way to categorize those trials as being congruent or incongruent.

2.2.2. Do numerical differences across the displays impact our CA tracking abilities?

2.2.2.1. Accuracy. Next, we examined the impact of numerical congruency. That is, relative to neutral trials (trials where the number of items was the same in both displays i.e., Number Ratio 1) did numerical congruency promote performance, did numerical incongruency detrimentally impact performance, or both? We performed a repeated measures ANOVA comparing performance across all three types of trials (neutral, congruent, incongruent; collapsing across all numerical and CA ratios). The main effect of congruency was significant, $F(2, 152) = 90.86$, $p < .001$, $\eta_p^2 = 0.55$. Paired samples *t*-tests revealed that performance on the incongruent trials (71.30%) was significantly worse than on the neutral trials (90.91%, $t(76) = 10.93$, $p < .001$, $d = 1.25$) and that performance on the congruent trials (93.55%) was significantly better than the neutral trials ($t(76) = 2.80$, $p < .01$, $d = 0.32$, See Fig. 3).

2.2.2.2. Reaction time. A repeated measures ANOVA comparing reaction times across neutral, congruent, and incongruent trials (collapsing across all numerical and CA ratios) also revealed a main effect of congruency, $F(2, 152) = 31.64$, $p < .001$, $\eta_p^2 = 0.30$. Similar to our findings with accuracy, a paired samples t-test revealed that reaction times on the congruent trials (968.02 s) were significantly faster than the neutral trials (1036.91 s, $t(75) = 5.35$, $p < .001$, $d = 0.62$), which, in turn, were significant faster than the incongruent trials (1074.19 s; $t(75) = 3.08$, $p < .01$, $d = 0.36$).

Thus, both accuracy and reaction time data suggest that conflicting numerical information (i.e., incongruency) detrimentally impacted performance relative to neutral trials, but consistent numerical information (i.e. congruency) also facilitated performance relative to neutral trials. Although incongruent numerical information appeared to detrimentally impact performance significantly more so than congruent numerical information benefited performance, it should be noted that the high level of performance on congruent trials may have led to ceiling effects in performance, limiting the extent to which performance could benefit from congruent numerical information.

2.2.3. How does set size impact CA acuity?

2.2.3.1. Accuracy. To investigate how differing set sizes impacted CA discrimination performance, we ran a repeated measures ANOVA comparing performance across the four set sizes (Single item, Small set, Medium set, Large set). Importantly because single item trials necessarily were number neutral (i.e., a comparison of one item to one item cannot involve congruent or incongruent trials) and because numerical congruency was unrelated to this question, we limited this analysis to only number neutral trials (those trials where number was identical in both arrays i.e., Number Ratio 1). Analyses revealed a significant effect of set size ($F(3,228) = 8.05$, $p < .001$, $\eta_p^2 = 0.10$), driven by significantly better performance on the Single trials (M = 94.36%) compared to Small (M = 91.22%), Medium (M = 90.26%), or Large sets (M = 89.81%, *p*'s < 0.001, *d*'s > 0.40). Although performance tended to decrease as a function of increasing set size, the difference in performance across Small, Medium, and Large set sizes did not reach significance (*p*'s > 0.15; See Fig. 4).

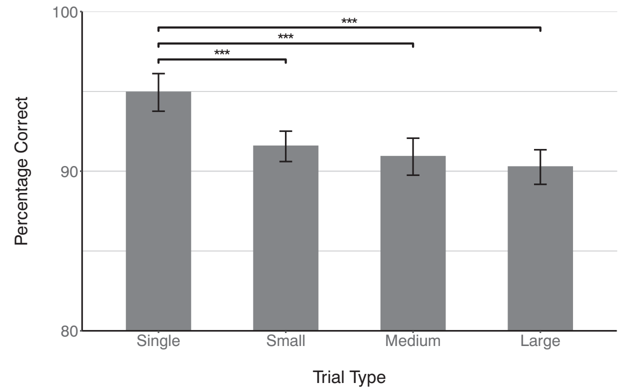

Fig. 4. Experiment 1: percent correct as a function of set size (single, small (4–7), medium (12–15) or large (20–25)). Error bars represent standard error. * $p < .05$, ** $p < .01$, *** $p < .001$

Because it is conceivable that the processes involved in tracking area in our single item trials may have been distinct from those involving more than one item (i.e., direct perception of area of a single item versus a potential computation process for tracking CA of a group of objects), we performed one additional analysis to explicitly compare accuracy as a function of set size for only those trials involving arrays of multiple items. We calculated the slope relating performance on small, medium, and large set sizes to a dummy variable (coding small as 1, medium as 2, and large as 3), examining only performance on the neutral trials, with a Number Ratio 1. We found a negative slope of −0.006 that did not differ significantly from 0, $t(77) = 1.38$, $p = .17$, $d = 0.15$.[6] This suggests that although performance decreased as set size increased, this trend was not significant.

2.2.3.2. Reaction time. A repeated measures ANOVA comparing reaction times across the four set sizes revealed a significant main effect ($F(3,222) = 11.68$, $p < .001$, $\eta_p^2 = 0.14$). Consistent with accuracy results, single item trials (M = 923.64 s) were processed significantly faster than trials involving arrays with multiple items (Small: M = 1067.04 s, Medium: M = 1025.68 s, Large: M = 1015.53 s; *p*'s < 0.001, *d*'s > 0.39). However, the pattern of results across sets containing more than one item was different. In particular, reaction times were significantly faster for large sets compared to small sets ($t(74) = 2.17$, $p = .03$, $d = 0.25$), although reactions times were similar for Small and Medium sets ($t(74) = 1.83$, $p = .07$, $d = 0.21$), and Medium and Large sets ($t(74) = 0.47$, $p = .64$, $d = 0.05$; See Fig. 5). We return to these results in the General Discussion.

We calculated the slope relating reaction times on small, medium, and large set sizes to a dummy variable and found that although the slope was negative (−20.68) it was not significantly different from 0, $t(74) = 1.83$, $p = .07$, $d = 0.21$.

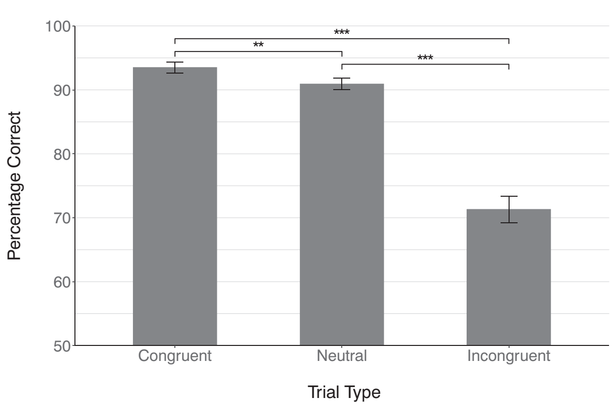

Fig. 3. Experiment 1: percent correct as a function of number ratio (1.33 and 1.5) and congruency (congruent and incongruent). Error bars represent standard error.
* $p < .05$, ** $p < .01$, *** $p < .001$

[6] Even when we excluded the easiest 1.9 ratio where participants were performing at ceiling, we found a negative slope of −0.005 that did not differ significantly from 0, $t(76) = 0.87$, $p = .39$, $d = -0.10$.

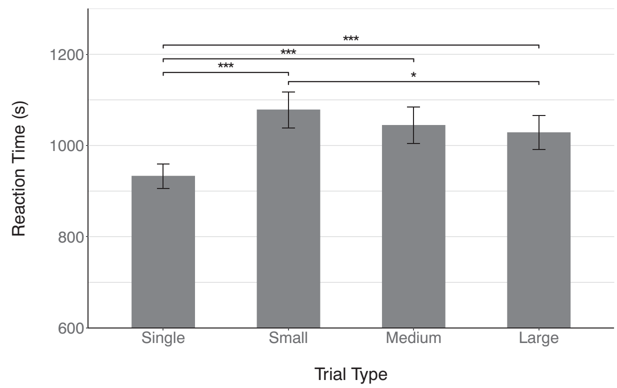

Fig. 5. Experiment 1: reaction time as a function of set size (single, small (4–7), medium (12–15) or large (20–25)). Error bars represent standard error. * $p < .05$, ** $p < .01$, *** $p < .001$

3. Experiment 2

Altogether, results from Experiment 1 suggest that CA discriminations in adults are ratio-dependent and hindered by numerical incongruency. Furthermore, we found that adults are significantly better (faster and more accurate) at discriminating the size of a single item compared to discriminating CA across multiple items. However, our results regarding how set size influences performance provided conflicting patterns of results across accuracy and RT, making it difficult to distinguish between our competing theories.

The goal of Experiment 2 was to replicate findings of Experiment 1 with three small changes. First, in Experiment 1, we controlled for *item* density by holding constant the number of items in the display per unit background area. Although this controlled for density of the number of items within the display, in retrospect, this did not control for the size of the items (and thus the inter-item distance) of the items within the display. That is, displays with greater cumulative area (holding number constant) necessarily had dots presented closer together in Experiment 1. Thus, in Experiment 2, we remedied this issue by controlling for *area* density by holding constant the cumulative surface area of the items per unit background to ensure that density did not drive our pattern of results. Second, in Experiment 2, we eliminated the 1.9 CA Ratio since adults performed near ceiling on this ratio, thus Experiment 2 tested adults on only 4 CA Ratios: 1.15, 1.3, 1.45 and 1.6. Lastly, we increased the number of trials participants completed in Experiment 2, providing greater confidence in our performance estimates.

3.1. Methods

The methods of Experiment 2 were identical to Experiment 1 except for the following:

3.1.1. Participants

A total of 54 undergraduate students from Boston College participated in our study in exchange for cash or course credit (41 female, M = 19.65 years, *Range* = 18–26 years). Since the effect sizes we obtained in Experiment 1 were larger than expected, we reduced our sample size in this experiment. All participants provided informed consent.

3.1.2. Procedure

Experiment 2 included the following variables and their levels: CA Ratio (1.15, 1.33, 1.45, 1.6), Number Ratio (1, 1.33 or 1.5), Congruency (Congruent, Incongruent), and Set Size (Small, Medium, Large). Similar to Experiment 1, we continued to have 38 trials per CA Ratio (this includes 2 single item trials per CA Ratio), leading to a total of 152 unique trials. To increase the precision in our measurement, we presented participants with the 152 unique trials 3 times over the course of the experiment (yielding 456 trials total). The trials were organized in blocks such that a participant was presented with all 152 unique trials before the trials would be repeated, with an unlimited break every 100 trials (5 breaks total). Otherwise, procedures were identical to Experiment 1.

3.1.3. Stimuli

The only changes made to the stimuli was that we now controlled for density by dividing the CA by the size of the display. The density was identical across the two displays to be compared in each trial, although across trials densities did vary from 0.08–0.15 (CA/background area in cm^2). Thus, the item background ranged from 193.3–300 cm^2.

3.2. Results & discussion

Consistent with the fact that we dropped the easiest CA ratio in this experiment, performance was significantly less accurate here compared to Experiment 1 ($t(116) = 2.19$, $p = .03$, $d = 0.40$), with an average of 80.19% correct (Range = 50.00–96.05%). Since we suspected that this was due to the exclusion of the easiest 1.9 CA Ratio, we performed a second independent samples *t*-test this time comparing performance on Experiment 1 and 2 excluding the easiest 1.9 CA Ratio in Experiment 1 as well (M = 83.47%, Range = 58.55–98.02%) and the difference in performance was no longer statistically significant ($t(116) = 1.03$, $p = .31$, $d = 0.19$), suggesting that lower performance in this Experiment was due to the fact that we eliminated our easiest CA Ratio.

The average weber fraction across participants was $w = 0.32$, (*Range* = 0.09–2.1, SD = 0.36, SE = 0.05), this was marginally worse than in Experiment 1, ($t(127) = 1.90$, $p = .06$, $d = 0.32$). We again used the Welch-Satterthwaite procedure for unequal variances to compare our weber fraction to that previously reported by Odic et al. (2013) and found a significant difference in performance ($t(54.96) = 3.43$, $p < .01$) suggesting that CA acuity in our study was significantly worse than prior reports of numerical acuity.

3.2.1. Does numerical congruency matter for CA judgments?

3.2.1.1. Accuracy. As in Experiment 1, we examined accuracy on trials involving arrays of multiple items in which number was either congruent or incongruent. In contradiction of predictions of the SoM Theory, a 2 (Number Ratio: 1.33 or 1.5) x 2 (Congruency: Congruent vs. Incongruent) repeated measures ANOVA again revealed a main effect of Congruency, ($F(1,53) = 60.62$, $p < .001$, $\eta_p^2 = 0.53$) such that participants performed significantly better on congruent (M = 91.70%) compared to incongruent (M = 62.90%) trials. We also replicated our main effect of Number Ratio ($F(1,53) = 14.19$, $p < .001$, $\eta_p^2 = 0.21$) with participants performing better on the 1.33 Ratio (M = 78.49%) compared to the 1.5 Ratio (M = 76.12%; See Fig. 6). However, unlike Experiment 1, we did not find a Number Ratio x Congruency interaction, $F(1,53) = 2.33$, $p = .13$, $\eta_p^2 = 0.04$ (although the pattern of results was identical across experiments).

3.2.1.2. Reaction time. The same 2 (Number Ratio: 1.33 or 1.5) x 2 (Congruency: Congruent vs. Incongruent) repeated measures ANOVA looking at participant reaction times again revealed a main effect of congruency ($F(1,51) = 21.24$, $p < .001$, $\eta_p^2 = 0.29$) with faster reaction times for congruent (M = 910.67 s) compared to incongruent trials (M = 1020.69 s). Similar to Experiment 1, reaction times for both Number Ratios were similar ($F(1,51) = 0.97$, $p = .33$, $\eta_p^2 = 0.02$) and we did not find a Number Ratio x Congruency interaction, $F(1,51) = 1.22$, $p = .27$, $\eta_p^2 = 0.02$.

3.2.2. Do numerical differences across the displays impact our CA tracking abilities?

3.2.2.1. Accuracy. As in Experiment 1, we examined to what extent congruency or incongruency between number and CA was helping or

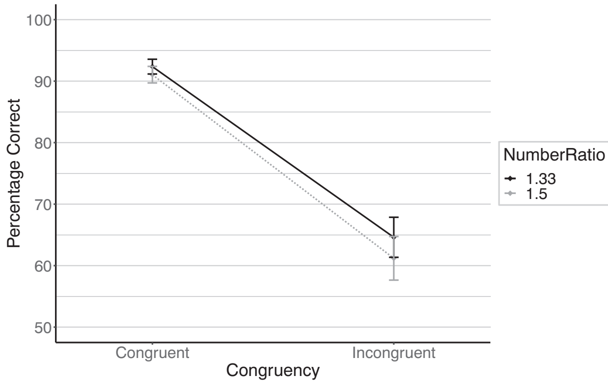

Fig. 6. Experiment 2: percent correct as a function of number ratio (1.33 and 1.5) and congruency (congruent and incongruent). Error bars represent standard error.

hurting performance in this task. Therefore, we ran a repeated measures ANOVA comparing performance on the number-neutral trials (trials of the Number Ratio 1) with performance of congruent trials and incongruent trials (again, collapsing all analyses across the 1.33 and 1.5 Number Ratio and across all CA ratios; See Fig. 7). Once again, the ANOVA was significant, $F(2, 106) = 61.29$, $p < .001$, $\eta_p^2 = 0.54$. Paired samples *t*-tests revealed that performance on congruent trials (91.72%) was significantly better than the neutral trials (84.28%, $t(53) = 4.82$, $p < .001$, $d = 0.66$), performance on the incongruent trials (62.90%) was significantly worse than the neutral trials (84.28%, $t(53) = 8.85$, $p < .001$, $d = 1.20$). Once again, this suggests that competing numerical information interferes with CA judgments, and consistent numerical information also facilitates CA judgments.

3.2.2.2. Reaction time. We conducted the same repeated measures ANOVA comparing reaction times across neutral, congruent, and incongruent trials which revealed a main effect of congruency, $F(2, 104) = 26.83$, $p < .001$, $\eta_p^2 = 0.34$. Although reaction times on congruent trials (934.45 s) were significantly faster than the neutral trials (1012.68 s, $t(52) = 6.67$, $p < .001$, $d = 0.92$), performance on the incongruent trials (1025.77 s) and neutral trials were similar ($t(52) = 1.17$, $p = .25$, $d = 0.16$).

3.2.3. How does set size impact CA acuity

3.2.3.1. Accuracy. To examine the effect of set size on CA discriminations, we again looked at those trials where number was held constant within trials but differed across trials. That is, we compared performance on Single trials to Small, Medium and Large trials involving the 1 Number Ratio. A within-subjects ANOVA again

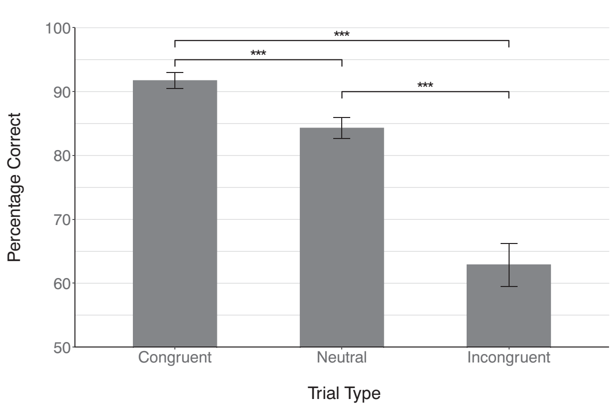

Fig. 7. Experiment 1: percent correct as a function of congruency (congruent, neutral and congruent). Error bars represent standard error.
†$p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

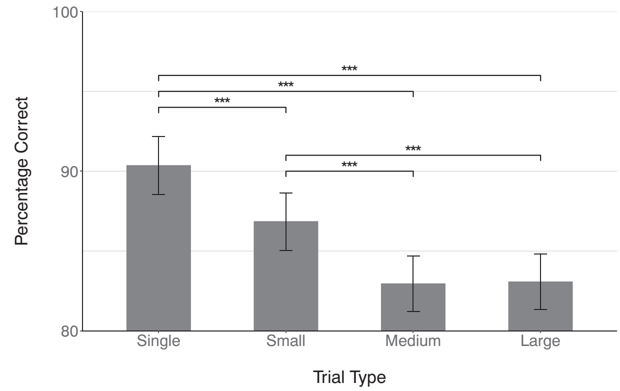

Fig. 8. Experiment 2: percent correct as a function of set size (single, small (2–4), medium (6–9) or large (9–15)). Error bars represent standard error.
* $p < .05$, ** $p < .01$, *** $p < .001$

revealed a significant effect of set size ($F(3,159) = 20.55$, $p < .001$, $\eta_p^2 = 0.28$; See Fig. 8), and follow-up paired samples t-tests revealed this was once again driven by the Single trials ($M = 90.35\%$) where participants performed significantly better than Small ($M = 86.84\%$), Medium ($M = 82.94\%$) or Large trials ($M = 83.06\%$; *p*'s < 0.001, *d*'s > 1.1). In contrast to Experiment 1, however, we found participants performed significantly better on Small compared to Medium and Large trials (*p*'s < 0.001).

Next, to specifically address the impact of set size on performance on trials involving sets, we calculated the slope for performance on small, medium, and large set sizes. Here we examined only performance on the neutral trials, with a Number Ratio 1, and found a negative slope (slope $= -0.019$) that differed significantly from 0 ($t(53) = 3.54$, $p < .001$, $d = -0.48$). This suggests that there was a steady decrease in performance as set size increased.

3.2.3.2. Reaction time. Next we examined this same question using reaction times. A repeated measures ANOVA comparing reaction times across the four set sizes revealed a significant main effect ($F(3,156) = 15.06$, $p < .001$, $\eta_p^2 = 0.14$). Consistent with previous findings, single item trials ($M = 863.76$ s) were processed significantly faster than trials involving arrays with multiple items (Small: $M = 1042.59$ s, Medium: $M = 1026.64$ s, Large: $M = 966.58$ s; *p*'s < 0.001, *d*'s > 0.51). Similar to our reaction time findings in Experiment 1, reaction times for Large sets were significantly faster than Small ($t(52) = 3.40$, $p < .001$, $d = 0.47$) and Medium sets ($t(52) = 2.67$, $p < .01$, $d = 0.37$), although reactions times were similar for Small and Medium sets ($t(52) = 0.50$, $p = .62$, $d = 0.07$; See Fig. 9). The slope of Small, Medium, and Large sets was negative (-20.68) and significantly different from 0, $t(52) = 3.40$, $p < .001$, $d = 0.47$.

Results from Experiment 2 replicate our findings in Experiment 1 revealing that CA discriminations abide by Weber's law and that congruency between number and CA plays an important role in discrimination performance. Once again, our results suggest that number is more likely to interfere with CA judgments than to assist them. Moreover, Set Size mattered in that adults were significantly better (faster and more accurate) at discriminating the size of single items compared to the CA of multiple items. Set size, however, impacted performance in distinct ways: larger sets were associated with lower accuracy but faster responding, indicative of a speed-accuracy trade-off as set sizes increased.

3.2.4. Combined analysis

3.2.4.1. Accuracy. We combined data from Experiment 1 and 2 and compared the average weber fraction across Experiments ($w = 0.27$, *Range* $= 0.07–2.1$), to weber fractions previously reported by Odic

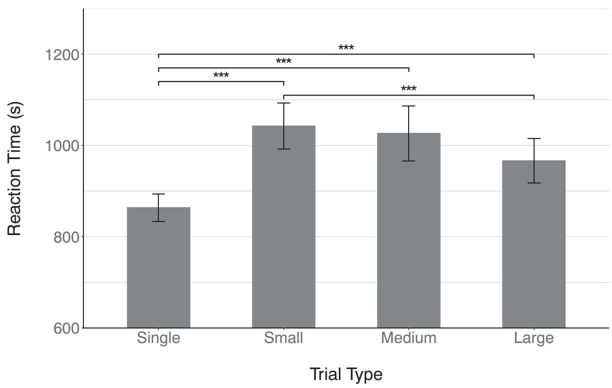

Fig. 9. Experiment 2: reaction time as a function of set size (single, small (2–4), medium (6–9) or large (9–15)). Error bars represent standard error.
* $p < .05$, ** $p < .01$, *** $p < .001$

et al. (2013) for adult numerical discrimination tasks (using the Welch-Satterthwaite procedure for unequal variances) and once again found a significant difference in performance, $t(23.17) = 4.09$, $p < .001$, with participants performing significantly worse on the CA task.

Given that our results regarding the effect of set size on accuracy were inconclusive (we found a significant effect of set size in Experiment 2 but not Experiment 1), we combined our data from Experiments 1 and 2 to run another slope analysis (excluding data from the 1.9 ratio in Experiment 1), again looking only at the neutral trials with Number Ratio 1. We found a negative slope of -0.011, which was significantly different from 0, $t(130) = 2.59$, $p = .011$, $d = -0.21$ suggesting that overall, participants did show increasingly worse performance as set size increased.

3.2.4.2. Reaction time. We also combined reaction times data for Experiment 1 and 2 and ran the same slope analysis. Here too the was negative -27.86 and significantly different from 0, $t(127) = 3.45$, $p < .001$, $d = 0.30$.

4. General discussion

The aim of this study was to investigate if and how CA discriminations are influenced by numerical information in adults. Although many previous studies have investigated adult abilities to discriminate discrete quantity (i.e. number discrimination; Halberda & Feigenson, 2008; Odic et al., 2013), very few studies have examined adults' performance on discrimination tasks that involve continuous properties, such as CA. This is a particularly interesting question given the claims made by proponents of the SoM theory that continuous quantities should be more easily represented than number because unlike number which is abstract, continuous quantities are perceptual in nature and are not tied to any specific sensory modality (Gebuis & Reynvoet, 2012b; Leibovich et al., 2017). As a direct test of these claims, many studies have investigated whether we can track number independent of continuous properties, finding that even human infants can do so (Halberda & Feigenson, 2008; Lipton & Spelke, 2003; Odic et al., 2013; Starr et al., 2013; Xu et al., 2005; Xu & Spelke, 2000). However, less work has explored the converse; that is, how well can we track continuous properties independent of number? (but see Barth, 2008; Hurewitz et al., 2006; Yousif & Keil, 2019).

Our first aim was to examine whether congruency between number and area played a role in adults' discrimination performance. If CA is significantly more salient and easy to represent than number, then CA acuity should be more precise than that of number. This did not appear to be the case. In addition to replicating previous findings suggesting that CA discriminations were ratio-dependent (Halberda & Feigenson, 2008; Odic et al., 2013), we also found that the average weber fraction associated with CA discriminations was significantly higher than those previously reported for number discrimination (Odic et al., 2013), contradicting any claims that adults are *better* at discriminating continuous quantities compared to number (Leibovich & Henik, 2014). Moreover, if CA is more salient and easy to represent than number, then numerical information should be less likely to interfere with CA representations than vice versa. Again, this did not appear to be the case. Across two experiments, we found that participants overwhelmingly performed better, and faster, on congruent trials (e.g. when the arrays with the larger number of dots also have a greater CA) compared to incongruent trials (e.g. when the larger number array has a smaller CA), replicating previous findings (Barth, 2008; DeWind & Brannon, 2012; Gebuis & Reynvoet, 2012a; Hurewitz et al., 2006). Not surprisingly, numerical information had a greater impact on adult performance for the most difficult CA judgments.

Moreover, our study expanded upon previous research by exploring whether numerical congruency facilitated CA judgments, whether numerical incongruency hindered CA judgments, or both, specifically when compared to neutral trials where the number of items in both arrays is identical (i.e. trials where the ratio of number was 1). We found both to be the case: incongruency between Number and CA significantly hurt performance *and* congruency between these two variables boosted performance, suggesting that adults *can* and *will* use all available quantity information in making quantitative judgments, whether or not this information is helpful or hurtful.

It should be noted that a recent paper has argued that numerical congruency may be better explained by a preference to judge Additive Area (AA; the sum of the length and width of the items in the array) instead of the true Cumulative Area of the displays (Yousif & Keil, 2019). Unfortunately, because AA nearly perfectly correlates with number in our experiments, it is impossible to determine whether number or AA provides a better account for our findings. However, there are many open questions regarding the AA theory (e.g., How do individuals track AA and why would such a tracking system develop? How does AA apply to real-world displays containing asymmetrical items?), and substantial research revealing number to be a relative and salient property of sets, making us cautious to interpret our findings in terms of AA at this time. Instead, we think that numerical congruency is the most likely account for the pattern seen in our data.

The final, and most important goal of this study was to understand the process by which adults represent CA when presented with an array of items. In particular, we compared two possible hypotheses. On the one hand, the 'Direct Perception' hypothesis assumed that we extract how much surface area we see in the display without any reliance upon individuating the items in the array, which is consistent with SoM theory (Gebuis & Reynvoet, 2012b; SoM: Leibovich et al., 2017; Mix et al., 2002). On the other hand, the 'Computation' hypothesis proposed that adults track the sizes of each individual item within the array and perform a summation process to arrive at an estimate of the CA of the array (as proposed by Barth, 2008). To distinguish between these two accounts, we investigated how set size affected performance on this task. Assuming that this summation process contributes error to the representation (Cordes et al., 2007), the 'Computation' hypothesis would predict a decrease in performance as the number of items in the array increased since each additional item adds to the error in the summation or computation process. The 'Direct Perception' hypothesis would not predict a relationship between CA acuity and set size.

On the one hand, both accuracy and RT analyses revealed that adults performed significantly more accurately and quicker on trials where they were presented with single items (trials that required them to make element area comparisons) compared to trials with small (4–7 dots), medium (12–15 dots), or large (20–25) sets. Although extreme predictions of the Direct Perception hypothesis may posit that single item trials would be tracked with the same precision and speed as trials involving multiple items, this clearly was not the case. Instead, it seems likely that the area of a single item is represented via a different process than that of the cumulative area of multiple items. In the former case, other continuous quantities – such as the diameter of the item – may serve as a more reliable cue for discrimination and thus area may not even be tracked under these circumstances. In the case of an array of items, though, it is unlikely that successful discrimination can take place without tracking the area of the items within the array (although see Yousif & Keil, 2019 for an alternative account).

Moreover, the fact that differences in acuity persist in judging the area of a single item versus an array of items emphasizes that if we do want to make comparisons between adult abilities to represent discrete and continuous properties, or make claims about the saliency of continuous variables in the context of numerical stimuli, it is important that we test both in the same context – that is, within arrays of objects. Given that tests of numerical discrimination by definition require adults to make estimates or computations across multiple items, one should similarly examine the representation of continuous dimensions in the context of sets, not just single items, in order to provide a fair comparison.

Our findings regarding the effect of set size on performance for trials involving multiple items, however, are much less clear. In the case of accuracy, slope analyses combining data from Experiments 1 and 2 demonstrated that as set size increased, accuracy in making CA judgments decreased. This accuracy finding provides support for the Computation Hypothesis, suggesting that when making CA judgments, adults represent the surface area of individual items within an array and sum across these representations to gain a representation of the array's CA. These findings are in line with previous findings by Barth (2008), whose computational model suggested that a computational account of CA representation provided the best explanation for the data.

On the other hand, RT analyses revealed an opposite pattern of results, such that participants took longer to respond when set sizes were small than when sets were large. This pattern was not predicted by either the Direct Perception or Computation hypothesis. Coupled with the pattern of results for accuracy, our findings suggest a speed-accuracy trade-off in terms of performance across set sizes. That is, when encountering small sets, participants may have felt more confident in their ability to judge the CA of the sets, and thus may have taken more time to provide more accurate judgments. In contrast, when presented with large sets of items, participants may have become discouraged and made a quick judgment that was less likely to be accurate. If so, our findings do suggest that set size plays a critical role in influencing the process of tracking CA, though perhaps not in the ways in which we originally hypothesized. Future research should further investigate CA judgments across differing set sizes to determine how confidence, and differing strategies, may vary across set sizes, and how these distinct strategies may provide insight into how CA is processed.

In conclusion, our results do not align with claims of a SoM theory. In particular, our results suggest that CA discriminations – in the context of multiple items – are not more precise than that of numerical discriminations, and may in fact be even *less* precise. Moreover, though it was completely irrelevant to the task demands and thus not a reliable cue for tracking, we find that number is automatically processed in the context of CA judgments suggesting that number is at least as salient as this continuous quantity. Most interestingly, our data produced opposing patterns of results when looking at the effects of set size on performance, such that responses were both less accurate and faster as set size increased, suggesting the implementation of distinct response strategies as set size increased. Future work should explore the impact of set size on cumulative area, and other continuous extent variables, more carefully to determine when and how adults invoke distinct strategies, and whether similar patterns are found across development.

Credit authorship contribution statement

Sophie Savelkouls: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Visualization, Writing - original draft, Project administration. **Sara Cordes:** Conceptualization, Methodology, Formal analysis, Supervision, Writing - review & editing, Resources, Funding acquisition.

References

Anderson, N. H., & Cuneo, D. O. (1978). The height + width rule in children's judgments of quantity. *Journal of Experimental Psychology: General, 107*(4), 335–378. https://doi.org/10.1037/0096-3445.107.4.335.

Barth, H. C. (2008). Judgments of discrete and continuous quantity: An illusory Stroop effect. *Cognition, 109*(2), 251–266. https://doi.org/10.1016/j.cognition.2008.09.002.

Boysen, S. T., Berntson, G. G., & Mukobi, K. L. (2001). Size matters: Impact of item size and quantity on array choice by chimpanzees (Pan troglodytes). *Journal of Comparative Psychology, 115*(1), 106–110. https://doi.org/10.1037/0735-7036.115.1.106.

Brannon, E. M., Abbott, S., & Lutz, D. J. (2004). Number bias for the discrimination of large visual sets in infancy. *Cognition, 93*(2), B59–B68. https://doi.org/10.1016/j.cognition.2004.01.004.

Brannon, E. M., Lutz, D., & Cordes, S. (2006). The development of area discrimination and its implications for number representation in infancy. *Developmental Science, 9*(6), https://doi.org/10.1111/j.1467-7687.2006.00530.x.

Brannon, E. M., & Terrace, H. S. (1998). Ordering of the numerosities 1 to 9 by monkeys. *Science, 282*(5389), 746–749. https://doi.org/10.1126/science.282.5389.746.

Cordes, S., & Brannon, E. M. (2008). The difficulties of representing continuous extent in infancy: Using number is just easier. *Child Development, 79*(2), 476–489. https://doi.org/10.1111/j.1467-8624.2007.01137.x.

Cordes, S., & Brannon, E. M. (2011). Attending to one of many: When infants are surprisingly poor at discriminating an item's size. *Frontiers in Psychology, 2*(APR), 65. https://doi.org/10.3389/fpsyg.2011.00065.

Cordes, S., Williams, C. L., & Meck, W. H. (2007). Common representations of abstract quantities. *Current Directions in Psychological Science, 16*(3), 156–161. https://doi.org/10.1111/j.1467-8721.2007.00495.x.

DeWind, N. K., & Brannon, E. M. (2012). Malleability of the approximate number system: Effects of feedback and training. *Frontiers in Human Neuroscience, 6*, 68.

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

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

Halberda, J., & Feigenson, L. (2008). Developmental change in the acuity of the "number sense": The approximate number system in 3-, 4-, 5-, and 6-year-olds and adults. *Developmental Psychology, 44*(5), 1457–1465. https://doi.org/10.1037/a0012682.

Hurewitz, F., Gelman, R., & Schnitzer, B. (2006). Sometimes area counts more than number. *Proceedings of the National Academy of Sciences, 103*(51), 19599–19604. https://doi.org/10.1073/pnas.0609485103.

Izard, V., Pica, P., Spelke, E. S., & Dehaene, S. (2008). Exact equality and successor function: Two key concepts on the path towards understanding exact numbers. *Philosophical Psychology, 21*(4), 491–505. https://doi.org/10.1080/09515080802285354.

Leibovich, T., & Henik, A. (2014). Comparing performance in discrete and continuous comparison tasks. *Quarterly Journal of Experimental Psychology, 67*(5), 899–917. https://doi.org/10.1080/17470218.2013.837940.

Leibovich, T., Henik, A., & Salti, M. (2015). Numerosity processing is context driven even in the subitizing range: An fMRI study. *Neuropsychologia, 77*, 137–147. https://doi.org/10.1016/j.neuropsychologia.2015.08.016.

Leibovich, T., Katzin, N., Harel, M., & Henik, A. (2017). From'sense of number'to'sense of magnitude'–the role of continuous magnitudes in numerical cognition. *Behavioral and Brain, 40*. https://doi.org/10.1017/S0140525X16000960.

Lidz, J., Pietroski, P., Halberda, J., & Hunter, T. (2011). Interface transparency and the psychosemantics of most. *Natural Language Semantics, 19*(3), 227–256. https://doi.org/10.1007/s11050-010-9062-6.

Lipton, J. S., & Spelke, E. S. (2003). Origins of number sense: Large-number discrimination in human infants. *Psychological Science, 14*(5), 396–401. https://doi.org/10.1111/1467-9280.01453.

Lourenco, S. F., & Bonny, J. W. (2017). Representations of numerical and non-numerical magnitude both contribute to mathematical competence in children. *Developmental Science, 20*(4), Article e12418. https://doi.org/10.1111/desc.12418.

Lourenco, S. F., Bonny, J. W., Fernandez, E. P., & Rao, S. (2012). Nonsymbolic number and cumulative area representations contribute shared and unique variance to symbolic math competence. *Proceedings of the National Academy of Sciences of the United States of America, 109*(46), 18737–18742. https://doi.org/10.1073/pnas.1207212109.

Marchant, A. P., Simons, D. J., & de Fockert, J. W. (2013). Ensemble representations: Effects of set size and item heterogeneity on average size perception. *Acta Psychologica, 142*(2), 245–250 Chicago.

McCrink, K., Dehaene, S., & Dehaene-Lambertz, G. (2007). Moving along the number line: Operational momentum in nonsymbolic arithmetic. *Perception & Psychophysics, 69*(8), 1324–1333. https://doi.org/10.3758/BF03192949.

Meck, W. H., & Church, R. M. (1983). A mode control model of counting and timing processes. *Journal of Experimental Psychology. Animal Behavior Processes, 9*(3), 320–334. https://doi.org/10.1037/0097-7403.9.3.320.

Mix, K. S., Huttenlocher, J., & Levine, S. C. (2002). *Math without words: Quantitative development in infancy and early childhood.* New York: Oxford University Press.

Moyer, R. S., & Bayer, R. H. (1976). Mental comparison and the symbolic distance effect. *Cognitive Psychology, 8*(2), 228–246. https://doi.org/10.1016/0010-0285(76)90025-6.
Odic, D. (2018). Children's intuitive sense of number develops independently of their perception of area, density, length, and time. *Developmental Science, 21*(2), Article e12533. https://doi.org/10.1111/desc.12533.
Odic, D., Le Corre, M., & Halberda, J. (2015). Children's mappings between number words and the approximate number system. *Cognition, 138*, 102–121. https://doi.org/10.1016/j.cognition.2015.01.008.
Odic, D., Libertus, M. E., Feigenson, L., & Halberda, J. (2013). Developmental change in the acuity of approximate number and area representations. *Developmental Psychology, 49*(6), 1103–1112. https://doi.org/10.1037/a0029472.
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. https://doi.org/10.3758/s13428-016-0772-7.
Savelkouls, S., & Cordes, S. (2017). Numerical intuitions in infancy: Give credit where credit is due. *Behavioral and Brain Sciences, 40*, Article e189. https://doi.org/10.1017/S0140525X16002272.
Starr, A., Libertus, M. E., & Brannon, E. M. (2013). Infants show ratio-dependent number discrimination regardless of set size. *Infancy, 18*(6), 927–941. https://doi.org/10.1111/infa.12008.
Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition, 74*(1), 1–11. https://doi.org/10.1016/S0010-0277(99)00066-9.
Xu, F., Spelke, E. S., & Goddard, S. (2005). Number sense in human infants. *Developmental Science, 8*, 88–101. https://doi.org/10.1111/j.1467-7687.2005.00395.x.
Yousif, S. R., & Keil, F. C. (2019). The additive-area heuristic: An efficient but illusory means of visual area approximation. *Psychological Science, 30*(4), 495–503. https://doi.org/10.1177/0956797619831617.