

REPORT

Heterogeneity impairs numerical matching but not numerical ordering in preschool children

Jessica Cantlon, Rebecca Fink, Kelley Safford and Elizabeth M. Brannon

Department of Psychology & Neuroscience, and Center for Cognitive Neuroscience, Duke University, USA

Abstract

Do preschool children appreciate numerical value as an abstract property of a set of objects? We tested the influence of stimulus features such as size, shape, and color on preschool children's developing nonverbal numerical abilities. Children between 3 and 5 years of age were tested on their ability to estimate number when the sizes, shapes, and colors of the elements in an array were varied (heterogeneous condition) versus when they did not vary (homogeneous condition). One group of children was tested on an ordinal task in which the goal was to select the smaller of two arrays while another group of children was tested on a match-to-sample task in which the goal was to choose one of two visual arrays that matched the sample in number. Children performed above chance on both homogeneous and heterogeneous stimuli in both tasks. However, while children showed no impairment on heterogeneous relative to homogeneous arrays in the ordering task, performance was impaired by heterogeneity in the matching task. We suggest that nonverbal numerical abstraction occurs early in development, but specific task objectives may prevent children from engaging in numerical abstraction.

Introduction

A hallmark of adult human numerical cognition is the ability to represent number across diverse arrays of discrete entities regardless of variability in perceptual features such as size, shape, or color. The principle governing this feature of adult numerical cognition has been termed the *abstraction principle* (Gelman & Gallistel, 1978; Zur & Gelman, 2004). The basic idea is that during enumeration, the counting algorithm ignores the identity of the to-be-counted elements such that enumeration of heterogeneous sets (e.g. a clover, a giraffe, and a jet) is accomplished just as readily as enumeration of homogeneous sets (e.g. three clovers). How does the abstraction principle develop? One possibility is that children's emerging number concepts already incorporate the abstraction principle. Under this scenario, the relative diversity of objects in a set would have no effect on accuracy in nonverbal numerical judgments. A second possibility, however, is that children's initial number concepts do not incorporate the abstraction principle and that only gradually, over development, do they come to understand that object identity has no impact on enumeration, whether verbal or nonverbal.

In the growing literature that documents the ability of preverbal infants to discriminate visual and auditory stimuli based on number, several studies have tested infants' nonverbal numerical discrimination of heterogeneous arrays and found no deficit relative to homogeneous arrays (e.g. Starkey, Spelke & Gelman, 1983; Strauss & Curtis, 1981; Feigenson, 2005). In fact, Feigenson (2005) suggested that heterogeneity *increased* the likelihood that infants would attend to number as opposed to continuous variables. Similarly, we recently reported that for rhesus monkeys, within-stimulus heterogeneity had no effect on accuracy in a numerical ordering task (Cantlon & Brannon, 2006; see also Church & Meck, 1984; Murofushi, 1997; Nieder, Freedman & Miller, 2002). If young infants and nonhuman animals can abstract across stimulus variability to represent number then we should also expect preschool children's early nonverbal numerical abilities to follow the abstraction principle. However, previous studies addressing the effect of heterogeneity on preschool children's numerical judgments have produced conflicting results.

A handful of studies have reported that preschool children perform better on a variety of numerical tasks with

Address for correspondence: Jessica Cantlon, Center for Cognitive Neuroscience, Duke University, Box 90999, Durham, NC 27708, USA; e-mail: jfc2@duke.edu

homogeneous arrays as compared to heterogeneous arrays (e.g. Von Gast, 1957; Siegel, 1974; Sophian, 2000; Mix, 1999; Mix, Huttenlocher & Levine, 1996). For example, Mix (1999) required different groups of children to nonverbally match arrays of small numbers of either heterogeneous or homogeneous objects. She found that 3–4-year-old children recognized numerical equivalence among homogeneous arrays but did not reliably recognize numerical equivalence among heterogeneous arrays. However, by 5 years of age, children accurately identified numerical equivalence for both homogeneous and heterogeneous arrays. In the domain of verbal counting, Von Gast (1957) required children to verbally identify the number of objects in an array and found that preschool children had great difficulty verbally labeling the sets of 1, 2, 3, 4, or 5 heterogeneous objects. Siegel (1974) also found that 4- and 5-year-old children showed impaired performance on a numerical matching task when the elements varied in shape and color as opposed to when they were identical.

To account for children's difficulty in abstracting numerical values from heterogeneous compared to homogeneous arrays, Mix (1999) proposed that the ability to represent number for heterogeneous arrays was dependent on the acquisition of number words. Children become better able to ignore superficial object features as they master the verbal counting system because number words embody abstract numerical categories. In support of this argument, there is evidence from non-numerical domains that children's conceptual abilities are facilitated by verbal labels (e.g. S. Gelman & Markman, 1986; Sandhofer & Smith, 2004).

However, at odds with the idea that language is necessary to make abstract numerical judgments is the finding that, under some circumstances, children as young as 3 years of age are equally able to represent numerical equivalence for heterogeneous and homogeneous arrays (Gelman & Tucker, 1975; see also Beckmann, 1924). Gelman and Tucker (1975) found that 3- and 4-year-old children who were trained to select a display with two or three homogeneous objects (two or three toy mice) as the 'winner' did not consider the display with a matching number of objects as the 'winner' if one of the objects in the 'winning' display (a toy mouse) was surreptitiously replaced with a physically different object (a toy soldier). But, when a separate group of 3- and 4-year-old children was trained in an identical paradigm with heterogeneous displays of two and three objects, these children successfully identified the display with a matching number of objects as the 'winner' even after the identity of one of the objects had been surreptitiously changed. Thus when trained with heterogeneous arrays, children were able to match based on quantity and ignore identity changes. From these results it seems that the characteristics of a numerical task can affect children's performance. Several

studies have further demonstrated the influence of contextual factors such as the linguistic demands of the task (e.g. Jordan, Huttenlocher & Levine, 1992, 1994) and the phrasing of task instructions (e.g. Gelman & Gallistel, 1978; Wynn, 1992; see Cowan, 1991, for review) on children's numerical performance. These observations suggest that children's performance on numerical tasks can be affected by variables orthogonal to numerical reasoning.

The conclusion that young children have difficulty enumerating heterogeneous arrays is largely based on findings from tasks that require children to match. One possibility is that children interpret the goal of a matching task as identifying overall similarity and this detracts from the salience of the numerical attribute of the stimuli.¹ The impaired performance on heterogeneous stimuli reported in previous studies could be better explained as a linguistic effect, related to young children's understanding of the task goals, as opposed to a conceptual effect of children's numerical reasoning. In fact, Cowan (1991) cites evidence that an adult understanding of the word 'same' develops gradually between 2 and 6 years of age and suggests that the gradual development of the meaning of 'same' affects young children's appreciation of numerical equivalence during this developmental period.

If poor performance on heterogeneous compared to homogeneous stimuli on nonverbal numerical matching tasks reflects delayed acquisition of the abstraction principle, then it should not be task specific. Conversely, if the stimulus effect is due to specific aspects of the matching task and the abstraction principle is already in effect in young children, then the effect of stimulus heterogeneity may not apply to other tasks. To test between these possibilities, we examined the influence of stimulus heterogeneity on preschool children's performance during two different nonverbal number tasks: a matching task and an ordinal task.

Method

General procedure

All children were tested at Duke University with approval from the Institutional Review Board. Eighty children were tested and were randomly assigned to the Ordinal or Matching Condition. Stimuli were presented by a RealBasic program on a *MagicTouch* screen (35 × 26 cm). To increase precision for recording children's response time (RT), children were required to initiate each trial by

¹ Consistent with this idea, many studies have shown that children fail to apply the appropriate conceptual focus in cognitive tasks even when they possess a given conceptual ability (e.g. Melkman & Deutsch, 1977; Gentner & Ratterman, 1991; Goswami & Brown, 1990).

touching a picture of a bunny presented in the bottom right-hand corner of the screen. The inter-trial interval (ITI) was thus participant controlled; however, it was a minimum of 1 second. Experimental sessions were on average 15–20 minutes.

Children were tested for a maximum of 64 trials in either condition and children who failed to complete at least 32 trials were excluded from the study. The overall percentage of children who were excluded from the study was 5%. Children who completed the minimum number of trials completed 54 trials on average in the Matching condition, and 59 trials on average in the Ordinal condition. Trials that children failed to complete were not included in our analyses. Children were given a series of post-test questions to assess their verbal counting ability.

Ordinal task

Participants

Participants were 40 children evenly divided into two age groups: 3- to 4-year-olds (Mean age = 3.5 years, range = 2.83–3.83) and 4- to 5-year-olds (Mean age = 4.7 years, range = 4.08–5.5). Of these children, 74% were from families in which both parents had college degrees and 26% of children had only one parent with a college degree. Eight-two percent of children were Caucasian and of the remaining children, 17% were African-American, 33% were of Hispanic descent, and 50% were of Asian descent. Data from one child were excluded for failing to complete the minimum number of trials.

Task and procedure

The task objective was to press the stimulus with the fewer number of elements. There were two numerosity pairs tested: 2 vs. 4 and 4 vs. 8. Correct responses were followed by brief positive visual and auditory computer feedback (1-second audio clip and picture of a sun). The experimenter also dropped a star sticker into a plastic cup for the child after each correct response. Incorrect responses caused the trial to terminate and were followed by a ‘Try Again’ audio clip and a 3-second black screen.

Instructions and demonstration

Prior to testing, the experimenter demonstrated the task to the child. Children were told that the task was to ‘touch the box with the smaller number of objects’. The experimenter demonstrated correct and incorrect trials. Children were encouraged to take their time to touch the bunny but to then respond as quickly as possible and not to count. If children counted aloud, they were interrupted and reminded not to

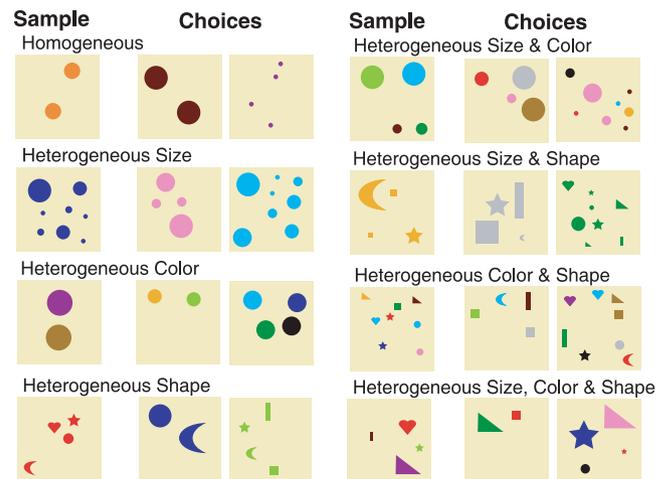


Figure 1 Exemplars of each of the eight stimulus conditions that were used for the Matching and Ordinal tasks.

Homogeneous condition: all elements within a stimulus were identical in size, shape, and color. Heterogeneous Color condition: elements were identical in size and shape, but varied in color. Heterogeneous Size condition: elements were identical in color and shape, but varied in size. Heterogeneous Shape condition: elements were identical in size and color but varied in shape. Heterogeneous Color & Size condition: elements were identical in shape, but varied in both size and color. Heterogeneous Color & Shape condition: elements were identical in size but varied in color and shape. Heterogeneous Size & Shape condition: elements were identical in color, but varied in size and shape. Heterogeneous Size, Color, & Shape condition: elements varied in size, color, and shape.

count. The numerical values used in demonstration trials were 2 and 8 and the elements within a stimulus were homogeneous but the size and color of the elements differed between stimuli. Once the child completed at least two trials correctly, he/she was allowed to begin the session.

Stimuli

Each stimulus was 6.2×6.2 cm and was randomly presented on the touch screen in one of six possible screen locations. Elements within each stimulus were randomly placed on a yellow stimulus background. Within-stimulus heterogeneity was systematically manipulated by allowing the elements within each stimulus to vary in size, shape, and/or color creating eight different stimulus conditions (Figure 1). There were four different exemplar pairs for each stimulus condition, two for the numerosity pair 2–4 and two for the numerosity pair 4–8. Thus, there were 32 different pairs of stimuli which were presented in random order.

Elements could be one of 15 different colors, five different sizes, and six different shapes (circles, stars, moons,

rectangles, triangles, and hearts). If the stimulus was homogeneous for a given dimension, a single value was randomly selected from the possible values.

To ensure that children could not use surface area to solve the task, the smaller numerosity had the greater cumulative surface area on half the trials for each numerosity pair in each of the eight stimulus conditions. Element sizes varied in diameter from 0.4 to 2.3 cm for both homogeneous and heterogeneous stimuli.

Matching task

Participants

Participants were 40 children who had not been tested in the Ordinal condition, and were evenly divided into two age groups: 3- to 4-year-olds (Mean age = 3.6 years, range = 3.08–4.0) and 4- to 5-year-olds (Mean age = 4.64, range = 4.04–5.92). Of these children, 67% were from families in which both parents had college degrees and 33% of children had only one parent with a college degree. Eighty-nine percent of children were Caucasian and of the remaining children, 25% were of Hispanic descent, and 75% were of Asian descent. Data from one child were excluded for failing to complete the minimum number of trials.

Task and procedure

On each trial, children were presented with a sample stimulus in the center of the screen that contained *N* elements. A press to the sample resulted in the presentation of two choice stimuli. The task was to press the stimulus with the same number of elements as the sample. The procedure was identical to that used in the Ordinal task.

Instructions and demonstration

The instructions and structure of demonstration was identical to the Ordinal task. However, in the Matching condition, the experimenter explained that the task was to 'touch the box with the same number of objects'. The sample value contained 2 elements and the choice stimuli contained 2 and 8 elements in demonstration trials. The choice stimuli were identical to those used in the demonstration of the Ordinal task.

Stimuli

The same 32 stimuli that were used as choice stimuli in the Ordinal task were used as choice stimuli in the Matching task and thus were from the same eight conditions shown in Figure 1. New stimuli were constructed to serve as samples. The numerical values of the choice

stimuli were always 2 vs. 4 and 4 vs. 8 and the sample contained 2, 4, or 8 elements.

The sample and choices were always from the same stimulus condition. The cumulative surface area of the sample stimulus was equal to the average cumulative surface area of the two choice stimuli, preventing subjects from using area as a basis for matching. The numerical values of the choice stimuli (2 vs. 4 and 4 vs. 8) occurred with equal probability. The distractor (incorrect choice) and correct choice always differed by a 1:2 ratio, thus the distractor for a sample of 2 was always 4, for 4 could be 2 or 8 and for 8 was always 4.

Verbal counting assays

Post-test, children were asked (1) How did you get so many answers right? (2) What was the rule in this game? (3) Do you remember how many objects you saw in the different boxes? Children were then tested on the 'How many?' and a simplified version of the 'Give a number' task following Wynn (1992). For the 'How many?' task, children were asked to verbally count six star stickers that were placed on the table. For the 'Give a number' task, children were asked to hand the experimenter six star stickers from a tin of foil stars. Children were asked to check their answers and corrections were recorded. If the child failed to correct a mistake, he/she was re-tested with a smaller number of objects until the child was able to accurately give *X* stars.

We used a scaling procedure for the verbal counting tasks similar to Mix (1999) with a range of 0–6. Children received 3 points if they successfully counted the six stars and told the experimenter how many stars there were on the 'How many?' task. Children received 2 points if they counted correctly but failed to report the total number of stars or if they counted the stars incorrectly but were able to correct their mistake. One point was given if a child failed to count the stars correctly but counted at least two stars. Children who counted in random sequences scored 0 points. On the 'Give a number' task, children received 3 points if they correctly counted six stars from the tin, they received 2 points if they produced an incorrect number of stars but were able to correct their mistake, and they received 1 point if they failed to correct their mistake but were able to count out at least two stars from the tin. Children who grabbed a random amount of stars or counted in random sequences scored 0 points.

Results

The main finding was that within-stimulus heterogeneity impaired accuracy on the Matching task and did not

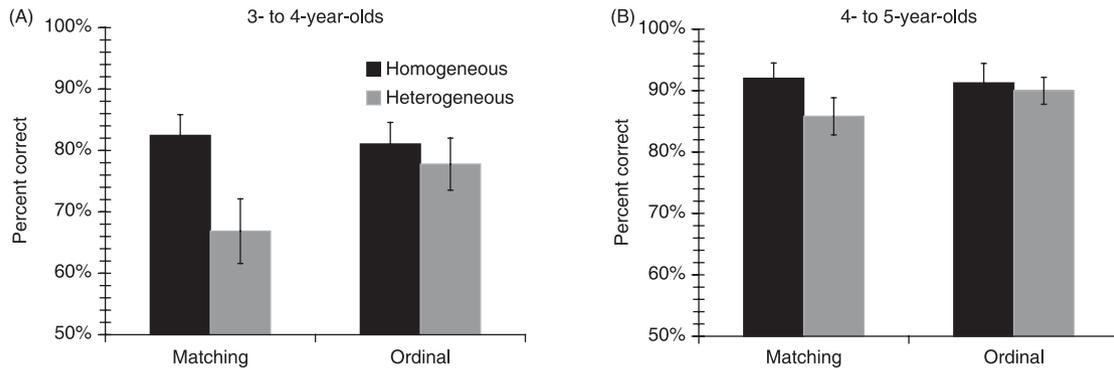


Figure 2 Accuracy on Homogeneous and Heterogeneous stimuli for 3- to 4-year-olds (A) and 4- to 5-year-olds (B), on the Matching task and Ordinal task. Children in both age groups showed higher accuracy on Homogeneous stimuli relative to Heterogeneous stimuli in the Matching task. In contrast, children showed equal accuracy on Homogeneous and Heterogeneous stimuli on the Ordinal task. There were no differences in overall accuracy between the two tasks for either age group. Chance = 50%.

Table 1 Mean accuracy and standard deviation for each stimulus condition by task and age group

	Ordinal		Matching	
	3–4 years	4–5 years	3–4 years	4–5 years
Homogeneous	81%	91%	82%	92%
Het size	79%	91%	60%	80%
Het color	74%	93%	62%	89%
Het shape	78%	89%	67%	92%
Het size color	71%	89%	77%	91%
Het size shape	82%	89%	64%	80%
Het color shape	81%	87%	64%	84%
Het size color shape	79%	93%	74%	85%

impair performance on the Ordinal task. A $2 \times 2 \times 2$ ANOVA for between-subject variables of Task (Matching or Ordinal) and Age (3–4 years or 4–5 years) and a within-subject variable of Stimulus Condition (Homogeneous or Heterogeneous) revealed a main effect of Age [$F(1, 76) = 15.14, p < .001$] and a significant interaction between Task and Stimulus Condition [$F(1, 76) = 10.53, p < .001$]. The main effect of Age was due to higher performance by 4- to 5-year olds on both tasks. Fisher's LSD post-hoc tests revealed that the interaction between Task and Stimulus Condition was due to significantly higher performance of both age groups on the homogeneous compared to the heterogeneous stimuli in the Matching task ($p < .0001$) and equal performance on homogeneous and heterogeneous stimuli for both age groups on the Ordinal task ($p = .23$). Children's accuracy on each of the eight stimulus conditions by age and task type are presented in Table 1 along with the standard deviation among children in each condition.

To investigate whether children showed improvement over the course of the session on specific stimulus classes, we compared accuracy for the first and second halves of the testing session. A Task (Matching or Ordinal) \times Block (First or Second Half) \times Stimulus Condition (eight conditions from Table 1) analysis revealed no main effect of Block [$F(1, 128) = 2.99, p = .09$] and, importantly, no interaction between Block and either of the two other variables [Task: $F(1, 128) = .29, p = .59$; Condition: $F(7, 896) = .54, p = .81$; Task and Condition: $F(7, 896) = 1.25, p = .27$]. Thus, children's performance on each task and stimulus condition appeared stable. Figure 2 shows children's accuracy on the homogeneous and heterogeneous stimulus conditions grouped by age and task type.

Comparison of overall accuracy between tasks

Both groups of children performed significantly above chance on both the Matching [chance = 50%; 3–4 years: Mean = 69%, $t(19) = 5.58, p < .01$; 4–5 years: Mean = 87%, $t(19) = 12.12, p < .01$] and Ordinal tasks [chance = .50; 3–4 years: Mean = 78%, $t(19) = 6.63, p < .01$; 4–5 years: Mean = 90%, $t(19) = 19.12, p < .01$]. There was no significant difference in accuracy on the Matching and Ordinal tasks for either age group [3–4 years: $t(38) = 1.74, p = .09$; 4–5 years: $t(38) = 1.00, p = .32$].² Children in both age groups performed significantly above chance on both tasks, regardless of variability in the cumulative surface area of the elements (all $ps < .001$).

² We decided to require all subjects in the ordinal task to choose the smaller of the two quantities because a pilot study with five subjects showed that accuracy was considerably higher when children were instructed to choose the larger quantity. The slightly lower accuracy for the 'choose smaller' version of the task was desirable in order to roughly equate accuracy between the Matching and Ordinal tasks.

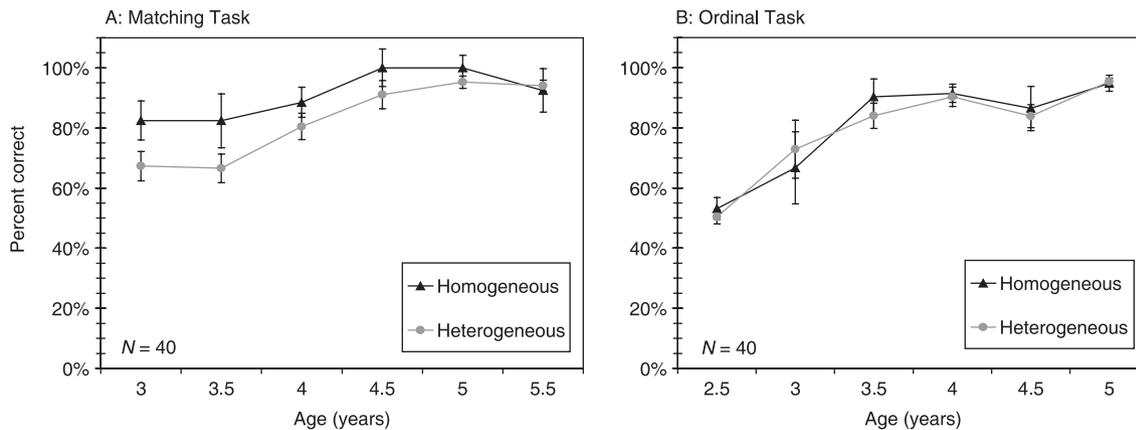


Figure 3 Accuracy on Homogeneous and Heterogeneous stimuli for children tested on the Matching task and Ordinal task. Children are grouped by age in half-year increments. The difference in accuracy between Homogeneous and Heterogeneous conditions gradually decreases with age on the Matching task but is equivalent for children on the Ordinal task from the earliest ages.

Comparison of accuracy between age groups

The 4- to 5-year-old children performed significantly better than the 3- to 4-year-old children on both tasks [Matching: $t(38) = 3.95$, $p < .01$; Ordinal: $t(38) = 2.55$, $p < .01$]. Children in both age groups performed significantly above chance on both the 2 vs. 4 [chance = 50%; 3–4 years: Mean = 77%, $t(39) = 9.40$, $p < .01$; 4–5 years: Mean = 91%, $t(39) = 29.74$, $p < .01$] and 4 vs. 8 [chance = 50%; 3–4 years: Mean = 71%, $t(39) = 6.57$, $p < .01$; 4–5 years: Mean = 90%, $t(39) = 24.05$, $p < .01$] numerical comparisons.

To better understand how these differences in performance on the Ordinal and Matching tasks unfold with development, we examined children's accuracy on each task and stimulus type by age in half-year increments (rounding down to the nearest .5 year), plotted in Figure 3. For the Ordinal task (Figure 3B), children between 3 and 3.5 years of age performed above chance and accuracy increased with age in parallel for homogeneous and heterogeneous stimuli. Accuracy increased with age equally for both homogeneous and heterogeneous stimuli until it approached ceiling at 5 years of age. On the Matching task (Figure 3A), children also performed above chance at 3–3.5 years of age; however, in contrast to the Ordinal task, children showed an early advantage in accuracy on homogeneous stimuli relative to heterogeneous stimuli that continued until approximately 5 years of age. Thus, the difference in accuracy between heterogeneous and homogeneous stimuli is highly correlated with age for the Matching task [Pearson product-moment; $r(39) = -.42$, $p < .01$] but not for the Ordinal task [Pearson product-moment; $r(39) = -.03$, $p = .83$].

Table 2 Comparison of response times (in milliseconds) by numerical pair for children on the Matching and Ordinal tasks

		2 vs. 4	4 vs. 8	$t(19)$	p
Ordinal	3–4 years	4402	4981	1.41	0.17
	4–5 years	1898	2124	2.74	<.05
Matching	3–4 years	4461	4767	0.86	0.4
	4–5 years	2670	3092	2.15	<.05

Parametric analysis of stimulus heterogeneity

The specific difficulties that younger children had with the heterogeneous stimuli on the Matching task are unclear. We found no significant effect of the number of dimensions varied on accuracy for the Matching task. A 2×3 ANOVA for a between-subject variable of Age (3–4 years or 4–5 years) and a within-subject variable of Number of Stimulus Dimensions varied (1, 2, or 3) revealed only a main effect of Age [$F(1, 38) = 18.64$, $p < .001$]. There was no significant effect of the Number of Dimensions Varied [$F(1, 38) = .51$, $p = .48$] and none of the interactions were significant.

Response time (RT) as a function of numerical size

Children typically took less than a half-second longer to respond when the numerical values were 4 and 8 compared to when they were 2 and 4. Table 2 shows the average RT to each number pair by task and age group and the results of the t -tests used to compare them.

Debriefing and verbal counting assays

Few children were able to report which numerical values were involved in the task. Forty-three percent of the 3- to 4-year-old children and 49% of the 4- to 5-year-old children reported seeing values such as 3, 6, 10, and 20 that were not presented in the task. Thirty-nine percent of the 3- to 4-year-olds and 13% of the 4- to 5-year-olds did not report any specific values.

On the verbal counting tasks, 45% of the 3- to 4-year-olds succeeded at the 'How many?' task with six objects and only 16% succeeded at the 'Give a number' task with six objects. Of the 4- to 5-year-old children, 75% succeeded at the 'How many?' task and 51% succeeded at the 'Give a number' task with six objects. Of the children who succeeded in the 'Give a number' task, 84% also succeeded in the 'How many?' task.

We used partial correlations to investigate the relationship between age, verbal counting ability, and the difference between performance on homogeneous and heterogeneous stimuli. As expected, counting ability was highly correlated with age [Pearson product-moment correlation; Matching task: $r = .61$, $p < .01$; Ordinal task: $r = .56$, $p < .01$]. Counting ability was also correlated with the difference in accuracy on homogeneous and heterogeneous stimuli for the Matching task [Pearson product-moment, $r = -.32$, $p < .05$] but not for the Ordinal task [Pearson product-moment, $r = .02$, $p = .86$]. To further examine the significant relationship between these two variables for the Matching task, we used a partial correlation to control for the factor of age. When age was partialled out, there was no longer a significant correlation between counting ability and the difference in accuracy between heterogeneous and homogeneous stimuli [Pearson product-moment, $r = -.09$, $p = .59$]. In contrast, when verbal counting ability was partialled out of the correlation, there was a marginally significant relationship between the difference in accuracy between homogeneous and heterogeneous stimuli and age [Pearson product-moment, $r = -.31$, $p = .05$]. Taken together, these analyses suggest that accuracy on both the homogeneous and heterogeneous conditions for the Matching task increases with age, independent of the development of verbal counting ability.

Discussion

The main finding of the present report is that homogeneity of elements within an array facilitated children's nonverbal numerical judgments in a task that required numerical matching but had no effect on children's ability to order numerical values. Previous studies of numerical matching in children have interpreted impaired perform-

ance on heterogeneous compared with homogeneous stimuli as evidence that young children do not represent number abstractly before they acquire number words (e.g. Mix, 1999). Our finding that young children have no difficulty nonverbally discriminating the numerical value of heterogeneous arrays in a numerical ordering task suggests that this conclusion must be revised.

This study was the first to test the effect of stimulus heterogeneity on children's numerical discriminations while parametrically varying the degree of within-stimulus heterogeneity. Although heterogeneity impaired accuracy on the Matching task, the degree of heterogeneity did not affect performance. One possibility is that large individual variability in the particular dimension(s) children find salient (e.g. color, size or shape) prevented a general parametric trend from emerging.

While we do not have a full explanation of why heterogeneity differentially impacts children's performance in the two tasks, we can offer some possibilities. First, it is important to note that the same stimuli were used in the Matching and Ordinal tasks and accuracy within each age group was equivalent between the two tasks, so differences in task difficulty could not have contributed to this pattern of results. We suggest that the type of experience children have with matching in their daily lives may bias them to attend to overall similarity or perceptual similarity in matching tasks. Children may have more experience identifying numerical relationships among perceptually similar objects than perceptually dissimilar objects. In general, statements linking homogeneous arrays of objects are more common than statements linking heterogeneous arrays of objects (Smith, 1993). Number words in particular are more commonly used to refer to sets of like objects ('There are three apples', 'There are three people', etc.) than sets of dissimilar objects ('There are three things'). Thus it is possible that children are better at counting, numerically labeling, or matching homogeneous compared with heterogeneous arrays simply because they are more likely to attend to number for homogeneous arrays, not because of an inability to apply the abstraction principle. When similarity is emphasized ('Choose the box with the *same* number of things'), children may attend further to overall similarity or similarity among task-irrelevant perceptual features like size, color, and shape (Cowan, 1991). In contrast, when children are instructed to choose the array with the *smaller* number of objects, as in the Ordinal task, there may be a narrower range of task-irrelevant features that children use in their comparisons.

Bauer and colleagues (Bauer & Mandler, 1989; Deak & Bauer, 1995, 1996) have shown that various features of a task can influence whether children are more or less likely to attend to perceptual similarity over more abstract

object relationships. In the same vein, our data demonstrate that the specific task a child engages in modulates the relative impact of perceptual similarity on their conceptual judgments. Specifically, when children make nonverbal numerical judgments, perceptual features are easily ignored in a task that emphasizes differences (Ordinal task) but prove to be problematic in a task that emphasizes similarity (Matching task). An interesting question is whether perceptual similarity has a differential effect on children's conceptual abilities in all tasks that emphasize likeness compared to those that emphasize difference. Within the domain of numerical cognition, several studies have demonstrated context effects on numerical performance in both verbal and nonverbal numerical tasks (e.g. Gelman & Gallistel, 1978; Gelman & Tucker, 1978; Jordan *et al.*, 1992, 1994; Wynn, 1992). Our results accord with these prior studies and provide an additional example of context effects on young children's nonverbal numerical reasoning.

A second important finding was that age and not verbal counting ability predicted the difference in accuracy on homogeneous and heterogeneous stimuli in the Matching task. Even though age and verbal counting ability were highly correlated, a partial regression determined that verbal counting ability did not explain any additional variance in the difference score between homogeneous and heterogeneous stimuli on the Matching task beyond the variance accounted for by age. These results contrast with those of Mix (1999), who reported that a child's mastery of the verbal counting system predicted his or her ability to nonverbally match heterogeneous arrays independent of age.

One possible explanation for the disparity between our results and those of Mix (1999) is that we used a different performance measure to examine the relationship between age, verbal counting ability, and numerical performance on heterogeneous stimuli. We assessed the relationship between age, verbal counting ability, and the *difference in accuracy* between homogeneous and heterogeneous stimuli, whereas Mix (1999) assessed the relationship between age, verbal counting ability, and *overall accuracy* on heterogeneous stimuli. Verbal counting ability may correlate with performance on heterogeneous stimuli because it correlates with general task competence; verbal counting ability may be an even better predictor of overall task competence than age. Our within-subjects assessment of children's relative performance on homogeneous and heterogeneous stimuli is likely a more sensitive measure of their numerical ability, independent of their overall task competence. On the basis of our assessment, it seems that verbal counting ability does not bear an important relationship to the development of nonverbal numerical abstraction. However, we note that

since all children in our study had *some* experience with verbal counting, we cannot rule out the possibility that a minimal amount of verbal counting experience is enough to instantiate the abstraction principle in nonverbal counting.

One might ask what type of numerical representation children employed in our two tasks. It is clear that children were not subitizing or using object-file representations in these tasks because both age groups were able to perform above chance expectations on comparisons of the values 4 and 8 which are values outside the range that can be handled by either process (Mandler & Sheebo, 1982; Trick & Pylyshyn, 1994). Multiple lines of evidence also indicate that children were not verbally counting in either task. First, the experimenter asked the children not to count and in the rare instance where a child attempted to verbally count aloud, they were interrupted and the instructions were repeated. Second, children's accuracy on the verbal counting tasks was not a significant correlate of their performance on either the Matching or Ordinal tasks, independent of age. Third, children's response times to the 4 vs. 8 numerical comparisons were on average only 400 milliseconds longer than to the 2 vs. 4 comparisons. If children were covertly counting, they should take at least twice as long to enumerate 12 (4 vs. 8) as opposed to 6 (2 vs. 4) elements, assuming equal counting time per object (Geary & Brown, 1991; Landauer, 1962). Finally, children were often unable to report any of the specific numerical values that appeared in the tasks.

Given that children were able to represent values as large as 4 or 8 and were unlikely to be verbally counting, it seems likely that they relied on the analog nonverbal system for representing number which has been hypothesized to underlie matching and ordinal numerical processes in animals, infants, children and even adult humans. An analog numerical system represents discrete numerical quantities as continuous magnitudes and obeys Weber's Law (e.g. Dehaene, 1997; Gelman & Gallistel, 2004; Gallistel & Gelman, 1992; Meck & Church, 1983); it is also hypothesized to enumerate entities independent of their identity or perceptual variability, and to develop early in the course of human development. Our finding that young children abstract numerical values from heterogeneous visual elements provides empirical support for these latter properties of the analog numerical model. A deeper investigation into young children's capacity for nonverbal numerical abstraction could examine their ability to abstract number from sets of entities that are not of one modality (see Church & Meck, 1984, for related data).

To sum up, previous findings of young children's impaired numerical performance on heterogeneous stimuli have been interpreted as evidence that numerical language

is necessary for the development of the abstraction principle. Although we successfully replicated the detrimental effect of heterogeneity in a nonverbal matching task (Mix, 1999), we found that young children were able to abstract number over heterogeneous stimuli when engaged in a numerical ordering task. Moreover, we found that verbal counting ability was not a good predictor of children's ability to represent number from heterogeneous arrays. We conclude that (1) children as young as 3 years of age can estimate the numerical value of a heterogeneous set of elements and (2) the abstraction principle for nonverbal counting is unlikely to develop from the acquisition of the verbal counting system. Thus *abstractness* may be an inherent property of young children's nonverbal numerical representation. What remains to be determined is whether children's capacity for numerical abstraction during nonverbal counting is a developmental precursor of numerical abstraction in verbal counting.

Acknowledgements

We thank Kerry Jordan, Evan MacLean, Mat Fleck, Umay Suanda and all members of the Brannon lab for help collecting and discussing these data. We thank Amy Needham for comments on an earlier version of this manuscript. This research was supported by an NSF Graduate Student Fellowship to JFC. This material is based upon work supported by the National Science Foundation under Grant No. 0448250 and a John Merck Fund fellowship to EMB.

References

- Bauer, P.J., & Mandler, J.M. (1989). Taxonomies and triads: conceptual organization in one- to two-year-olds. *Cognitive Psychology*, **21** (2), 156–184.
- Beckmann, H. (1924). Die Entwicklung Der Zahlleistung Bei 2–6 Jahrigen Kindern. *Zeitschrift fur Angewandte Psychologie*, **22**, 1–72.
- Cantlon, J.F., & Brannon, E.M. (2006). The effect of heterogeneity on numerical ordering in rhesus monkeys. *Infancy*, **9** (2), 173–189.
- Church, R., & Meck, W. (1984). The numerical attribute of stimuli. In H.L. Roitblat, T.G. Bever, & H.S. Terrace (Eds.), *Animal cognition* (pp. 445–464). Hillsdale, NJ: Erlbaum.
- Cowan, R. (1991). The same number. In K. Durkin & B. Shire (Eds.), *Language in mathematical education: Research and practice* (pp. 48–58). Milton Keynes: Open University Press.
- Deak, G., & Bauer, P.J. (1995). The effects of task comprehension on preschoolers' and adults' categorization choices. *Journal of Experimental Child Psychology*, **60** (3), 393–427.
- Deak, G.O., & Bauer, P.J. (1996). The dynamics of preschoolers' categorization choices. *Child Development*, **67** (3), 740–767.
- Dehaene, S. (1997). *The number sense: How the mind creates mathematics*. New York: Oxford University Press.
- Feigenson, L. (2005). A double-dissociation in infants' representation of object arrays. *Cognition*, **95**, B37–B48.
- Gallistel, C.R., & Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition*, **44** (1–2), 43–74.
- Geary, D.C., & Brown, S.C. (1991). Cognitive addition: strategy choice and speed-of-processing differences in gifted, normal, and mathematically disabled children. *Developmental Psychology*, **27** (3), 398–406.
- Gelman, R., & Gallistel, C.R. (1978). *The child's concept of number*. Cambridge, MA: Harvard University Press.
- Gelman, R., & Gallistel, C.R. (2004). Language and the origin of numerical concepts. *Science*, **306** (5695), 441–443.
- Gelman, R., & Tucker, M.F. (1975). Further investigations of the young child's conception of number. *Child Development*, **46**, 167–175.
- Gelman, S.A., & Markman, E.M. (1986). Categories and induction in young children. *Cognition*, **23** (3), 183–209.
- Gentner, D., & Rattermann, M.J. (1991). Language and the career of similarity. In S.A. Gelman & J.P. Byrnes (Eds.), *Perspectives on thought and language: Inter-relations in development* (pp. 225–277). London: Cambridge University Press.
- Goswami, U., & Brown, A.L. (1990). Higher-order structure and relational reasoning: contrasting analogical and thematic relations. *Cognition*, **36** (3), 207–226.
- Jordan, N., Huttenlocher, J., & Levine, S.C. (1992). Differential calculation abilities in young children from middle and low income families. *Developmental Psychology*, **28**, 644–653.
- Jordan, N., Huttenlocher, J., & Levine, S.C. (1994). Assessing early arithmetic abilities: effects of verbal and nonverbal response types on the calculation performance of middle and low income children. *Learning and Individual Differences*, **6**, 413–442.
- Landauer, T.K. (1962). Rate of implicit speech. *Perceptual and Motor Skills*, **15**, 646.
- Mandler, G., & Shebo, B. (1982). Subitizing: an analysis of its component processes. *Journal of Experimental Psychology: General*, **111**, 1–22.
- 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.
- Melkman, R., & Deutsch, H. (1977). Memory functioning as related to developmental changes in bases of organization. *Journal of Experimental Child Psychology*, **23**, 84–97.
- Mix, K. (1999). Similarity and numerical equivalence: appearances count. *Cognitive Development*, **14** (2), 269–297.
- Mix, K.S., Huttenlocher, J., & Levine, S.C. (1996). Do preschool children recognize auditory–visual numerical correspondences? *Child Development*, **67**, 1592–1608.
- Murofushi, K. (1997). Numerical matching behavior by a chimpanzee (*Pan Troglodytes*): subitizing and analogue magnitude estimation. *Japanese Psychological Research*, **39** (3), 140–153.
- Nieder, A., Freedman, D., & Miller, E. (2002). Representation of the quantity of visual items in the primate prefrontal cortex. *Science*, **297** (5587), 1708–1711.

- Sandhofer, C.M., & Smith, L.B. (2004). Perceptual complexity and form class cues in novel word extension tasks: how 4-year-old children interpret adjectives and count nouns. *Developmental Science*, **7** (3), 378–388.
- Siegel, L.S. (1974). Heterogeneity and spatial factors as determinants of numeration ability. *Child Development*, **45**, 532–534.
- Smith, L.B. (1993). The concept of same. In Hayne W. Reese (Ed.), *Advances in child development and behavior* (vol. 24; pp. 215–252). San Diego, CA: Academic Press.
- Sophian, C. (2000). From objects to quantities: developments in preschool children's judgments about aggregate amount. *Developmental Psychology*, **36**, 724–730.
- Starkey, P., Spelke, E., & Gelman, R. (1983). Detection of intermodal numerical correspondences by human infants. *Science*, **222** (4620), 179–181.
- Strauss, M., & Curtis, L. (1981). Infant perception of numerosity. *Child Development*, **52**, 1146–1152.
- Trick, L.M., & Pylyshyn, Z.W. (1994). Why are small and large numbers enumerated differently? A limited-capacity pre-attentive stage in visions. *Psychological Review*, **101** (1), 80–102.
- Von Gast, H. (1957). Der Umgang mit Zahlen und Zahlgebilden in der frühen Kindheit. *Zeitschrift für Psychologie*, **161**, 1–90.
- Wynn, K. (1992). Childrens acquisition of the number words and the counting system. *Cognitive Psychology*, **24** (2), 220–251.
- Zur, O., & Gelman, R. (2004). Young children can add and subtract by predicting and checking. *Early Childhood Research Quarterly*, **19**, 121–137.

Received: 11 May 2006

Accepted: 2 August 2006