The role of pattern recognition in children’s exact enumeration of small numbers

Brenda R. J. Jansen1*, Abe D. Hofman2, Marthe Straatemeier2, Bianca M. C. W. van Bers1, Maartje E. J. Raijmakers1 and Han L. J. van der Maas2

1Department of Psychology, Developmental Psychology, University of Amsterdam, The Netherlands
2Department of Psychology, Psychological Methods, University of Amsterdam, The Netherlands

Enumeration can be accomplished by subitizing, counting, estimation, and combinations of these processes. We investigated whether the dissociation between subitizing and counting can be observed in 4- to 6-year-olds and studied whether the maximum number of elements that can be subitized changes with age. To detect a dissociation between subitizing and counting, it is tested whether task manipulations have different effects in the subitizing than in the counting range. Task manipulations concerned duration of presentation of elements (limited, unlimited) and configuration of elements (random, line, dice). In Study 1, forty-nine 4- and 5-year-olds were tested with a computerized enumeration task. Study 2 concerned data from 4-, 5-, and 6-year-olds, collected with Math Garden, a computer-adaptive application to practice math. Both task manipulations affected performance in the counting, but not the subitizing range, supporting the conclusion that children use two distinct enumeration processes in the two ranges. In all age groups, the maximum number of elements that could be subitized was three. The strong effect of configuration of elements suggests that subitizing might be based on a general ability of pattern recognition.

Subitizing, the ability to rapidly and accurately enumerate a small set of elements (Kaufman, Lord, Reese, & Volkman, 1949), is a component of number sense, which is essential for proficient math performance (Jordan, Kaplan, Locuniak, & Ramineni, 2007; Kroesbergen, van Luit, van Lieshout, van Loosbroek, & van de Rijt, 2009). Deficient subitizing is suggested to underlie lagging math skills of children with dyscalculia (Schleifer & Landerl, 2011). Despite extensive work on subitizing (starting with Kaufman et al., 1949), the question whether subitizing is a separate process, dissociable from estimation and counting is still actively investigated in various domains, such as neuropsychology (e.g., Dehaene & Cohen, 1994; Demeyere, Lestou, & Humphreys, 2010; Harvey, Klein, Petridou, & Dumoulin, 2013; Nan, Knösche, & Luo, 2006), psychonomics (Watson, Maylor, & Bruce, 2007), and developmental psychology (e.g., Schleifer &

*Correspondence should be addressed to Brenda R. J. Jansen, Weesperplein 4, 1018 XA Amsterdam, The Netherlands (email: b.r.j.jansen@uva.nl).

DOI:10.1111/bjdp.12032
In this article, we use a new method to test whether subitizing is a separate process. This new method also allows us to investigate the relation between subitizing and pattern recognition.

The standard method to differentiate between subitizing and counting involves a bilinear model, which statistically describes the shape of the function relating response times (RTs) to numerosity. It combines a regression function with a small positive, near-zero slope for the subitizing range; and a regression function with a larger, positive slope for the counting range. Reeve, Reynolds, Humberstone, and Butterworth (2012) additionally represented the dissociation between subitizing and counting by a change from a linear to an exponential function. The transition point is often set between 3 and 4 (e.g., Akin & Chase, 1978; Chi & Klahr, 1975; Mandler & Shebo, 1982; van Oeffelen & Fox, 1982; Trick & Pylyshyn, 1993) but is closer to 3 for children (Maylor, Watson, & Hartley, 2011; Svenson & Sjöberg, 1983).

The present study uses an alternative method, put forward by Trick (2008), to investigate the existence of a distinction in preschoolers’ enumeration skills and a possible development of the subitizing range. The procedure circumvents two problems of the standard method. First, individual differences in the transition point can only be detected by estimating statistical models per participant (see Balakrishnan & Ashby, 1992; Plaisier, Bergmann Tiest, & Kappers, 2009, 2010). This requires many administrations of each numerosity to each participant, which is unfeasible with children. Also, simulations demonstrate that modern techniques (e.g., Muggeo, 2003) often fail in detecting transition points and differences in slope, when ranges of numerosities are small and slopes of the regression models differ in size, but not in sign (Julious, 2001; Muggeo, 2003). Second, good fit measures that allow model comparison are lacking. Trick argues that the hypothesis of a distinction is supported if manipulation of task conditions changes performance in the ranges differently. For example, using differently coloured elements speeds enumeration in the counting range, but not the subitizing range.

This study is the first to apply Trick’s procedure with preschoolers. It is remarkable that developmental studies of enumeration are scarce (but see Benoit, Lehalle, & Jouen, 2004; Starkey & Cooper, 1995) because enumeration processes develop in preschoolers and not in adults. We present developmental data from both an experimental (Study 1) and a field study (Study 2). Study 1 includes a controlled manipulation of task conditions in a selected (small) sample, whereas Study 2 is less controlled, but has a very large sample, from a wide background, with many repetitions per participant.

Time limit was manipulated in Study 1. Performance in a condition with limited presentation duration was compared to performance in an unlimited time condition. Starkey and Cooper (1995) conclude that performance of children from 2 to 5 years of age is accurate in the subitizing but not the counting range in a limited time condition. However, only 2-year-olds’ performance was compared in two time conditions. If subitizing and counting are the same process, manipulation of time limit is expected to affect performance comparably in the subitizing and the counting range. If subitizing is a separate process, variation in time limit is expected not to affect performance in the subitizing range because subitizing would ensure high performance. Limited time would deteriorate performance in the counting range as counting would be impossible and children probably resort to estimation.

Configuration of the elements to be enumerated was manipulated in both Study 1 and Study 2, using random, dice, and line configurations. The use of these various configurations is theoretically important because it allows studying the nature of
subitizing, which is thought to be related to pattern recognition (Mandler & Shebo, 1982). Dice configurations form familiar patterns that may be processed holistically. Enumeration performance improves when presenting familiar configurations (e.g., four dots presented as vertices of a square) as compared to random configurations (e.g., Benoit et al., 2004). As random, dice, and line configurations differ minimally for numbers in the subitizing range, it is expected that configuration would not affect performance in the subitizing range. Presenting elements in dice patterns is expected to benefit performance in the counting range. Line configurations also form a pattern ('line'), but it does not relate to a specific number of elements. Presenting elements in a line may either facilitate performance in the counting range, because elements are easily detected, or complicate performance because one may easily skip or recount elements (Towse & Hitch, 1996). The age variety in both studies, combined with the manipulation of configurations, allows for studying development of the subitizing range when presented with various configurations.

Manipulating the configuration of the elements is particularly important for investigating the relation between subitizing and pattern recognition. If familiarity of configurations has a larger effect on performance than the number of elements in the configuration, this is an important indication that pattern recognition is central in subitizing.

STUDY 1
Method
Participants
Participants came from two Dutch primary schools. Socio-economical status was high in one school and low in the other. Parents either signed informed consent to allow their child’s participation (high SES-school) or were informed and could refuse participation (low SES-school). The Local Ethics committee approved of the procedures. A total of twenty-six 4-year-olds (58% girls) and thirty-seven 5-year-olds (51% girls) participated.

The final sample consisted of nineteen 4-year-olds ($M = 4.59$ years, $SD = 0.24$, 47% girls) and thirty 5-year-olds ($M = 5.43$ years, $SD = 0.33$, 53% girls), who completed at least 75% of the problems in both versions. Children missed problems due to inattentiveness. Inclusion was independent of age group, $\chi^2(1) = 0.57, p = .452$, gender, $\chi^2(1) = 0.77, p = .380$, and condition, $\chi^2(2) = 3.93, p = .140$.

Material
Task presentation was on 15-inch laptops, with a screen resolution of $1,024 \times 768$ pixels. The screen was viewed from a distance of about 40 cm. Red dots (RGB-values: 255, 0, 0; diameter: 1 cm; $1.4^\circ$) were presented in a screen-centred black-bordered white square ($10 \times 10$ cm; $14^\circ \times 14^\circ$). The minimum interdot distance was $1.79^\circ$. The square covered an $8 \times 8$ cm ($11^\circ \times 11^\circ$) matrix. Response buttons representing numbers 1–7 and a question mark were displayed at the bottom of the screen.

All subjects performed a task version with presentation duration limited to 250 ms and a task version with unlimited presentation duration. Order of versions varied randomly. Four consecutive screens appeared: (1) fixation cross (500 ms), (2) presentation of dots (250 ms in limited time version; user-terminated in unlimited time version), (3) mask
Two general example problems and two version-specific examples were part of the task. Test problems were presented in four blocks of six problems each, yielding 24 problems. In each block, displays of 1–6 elements were presented randomly. Hence, each numerosity (1–6) was presented four times. Finally, 10 dots were presented in a line on screen.

Participants were randomly placed in the random, line, or dice condition. In the random condition, elements were spread in an unsystematic way. Four different displays were used for each number. In the line condition, evenly spaced elements were aligned on a horizontal or vertical line. Two different horizontal and two different vertical configurations were used for each number, varying distance between elements and length of complete display. In the dice condition, elements were presented in a dice pattern. The same dice configuration was used for each number, but absolute distance between elements varied, thereby varying the size of the entire display between displays of the same numerosity. Position of the display in the white square varied. Figure 1 shows examples of each configuration.

Figure 1. Study 1: Enumeration problems with three, four, and five elements (from left to right) in three types of configurations. From top to bottom: Random, dice, and line.
**Procedure**

Individual task administration took place in a quiet room at school. Instructions were printed on screen and read out loud by the experimenter. She told that ‘grand dad collected berries for a bird near his house’ and announced that berries would be shown after presentation of a small cross. At presentation of the elements, she asked: ‘Can you tell me how many berries grand dad collected? Tell me how many and I will click on the answer’. Dots were presented until the child answered or indicated he/she did not know the answer. Correct responses varied from 1 to 6, whereas response buttons indicated numbers 1–7. The experimenter clicked on the button corresponding to the stated number or on the question mark in the absence of a response. The second example followed.

The experimenter introduced the two versions by encouraging the child to pay extra attention in the limited time condition and explaining that the berries would only disappear if the child pressed the space bar in the unlimited time condition.

In the final task, the experimenter asked ‘Can you count these dots for me?’ and pointed to the 10 dots on screen. Time was unlimited. The experimenter noted whether the child was able to count up to six or made any errors. We assured that counting skills were sufficient for counting to six so that performance variations could be attributed to experimental manipulations.

**Results**

Multivariate ANOVA with error rates (number of errors expressed as a proportion) in the unlimited and limited time condition as dependent variables and order and configuration as between-factors indicated no main effect of order, $F(11, 33) = 0.99, p = .476$, and no interaction effect between order and configuration, $F(22, 68) = 1.34, p = .181$. Hence, data from different orders were combined.

A mixed ANOVA with between-subjects factor configuration (three levels: Random, dice, line) and age (two levels: 4 and 5 years) and within-subjects factors duration (two levels: Limited and unlimited) and numerosity (six levels: 1, 2, 3, 4, 5 or 6 elements) was performed on the error rates. In fact, displays with one element were similar in the three configurations.

The main effect of configuration was significant, $F(2, 43) = 5.47, p = .008, \eta^2_p = .20$. Post-hoc analyses using Tukey’s HSD indicated that error rates of random and line presentations were significantly higher than those of dice presentations. The main effect of age was not significant, $F(1, 43) = 2.78, p = .103$. The main effect of duration was significant, $F(1, 43) = 75.79, p < .001, \eta^2_p = .64$, with lower error rates in the unlimited than the limited time condition. Finally, the main effect of numerosity was significant, $F(5, 215) = 56.19, p < .001, \eta^2_p = .57$. Post-hoc analyses demonstrated that error rates increased with every additional element, except for the increase from 1 to 2 elements and the increase from 5 to 6 elements.

However, main effects were qualified by interactions between configuration and duration, $F(2, 43) = 7.84, p = .001, \eta^2_p = .27$, configuration and numerosity, $F(10, 215) = 5.27, p < .001, \eta^2_p = .20$, and duration and numerosity, $F(5, 215) = 13.56, p < .001, \eta^2_p = .24$, and three-way interactions between configuration, duration, and numerosity, $F(10, 215) = 2.54, p = .006, \eta^2_p = .11$, and between configuration, age, and numerosity, $F(10, 215) = 2.07, p = .028, \eta^2_p = .088$. Remaining two-, three-, and four-way interactions were not significant.
ANOVA\s with factor configuration were performed by numerosity and by duration to further investigate the interaction between configuration, duration, and numerosity. Alpha level was divided by the number of ANOVA\s performed for each duration condition (6) and adjusted to .008. In the limited time condition, error rates did not differ between configurations for one, two, and three elements, $F(2, 46) = 0.01$, $p = .990$; $F(2, 46) = 1.28$, $p = .289$; $F(2, 46) = 0.79$, $p = .460$, for 1–3 elements, respectively. The error rates for four elements just did not differ between configurations, due to the corrected alpha level, $F(2, 46) = 4.13$, $p = .022$. Error rates did differ between configurations for five elements, $F(2, 46) = 12.57$, $p < .000$, $\eta_p^2 = .55$, and six elements, $F(2, 46) = 7.98$, $p = .001$, $\eta_p^2 = .35$. Post-hoc analyses showed that error rates were higher for random and line patterns than for dice patterns and for both five and six elements. Error rates did not differ between configurations for any numerosity in the unlimited time condition, $F(2, 46) = 0$; $F(2, 46) = 1.13$, $p = .332$; $F(2, 46) = 1.75$, $p = .185$; $F(2, 46) = 0.47$, $p = .630$; $F(2, 46) = 4.26$, $p = .020$; $F(2, 46) = 0.25$, $p = .782$, for 1–6 elements, respectively.\(^1\) Note that the configuration effect just did not reach significance for enumerating five elements in the unlimited time condition, due to the corrected alpha level.

The three-way interaction was also investigated by testing the duration effect, by numerosity, for random configurations only, performing paired-samples $t$-tests. Alpha level was adjusted to .008. Error rates of enumerating one and two elements did not significantly differ between the two duration conditions, $t(13) = 1.47$, $p = .165$ for one element; $t(15) = 1.00$, $p = .336$ for two elements. The duration effect for enumerating three elements just did not reach significance, due to the corrected alpha level, $t(13) = 2.69$, $p = .019$. In contrast, error rates for enumerating four, five, and six elements were higher in the limited compared with the unlimited time condition, $t(13) = 5.70$, $p < .001$, $r = .85$; $t(13) = 4.94$, $p < .001$, $r = .81$; $t(13) = 5.95$, $p < .001$, $r = .86$ for four, five, and six elements.\(^2\)

Repeated-measures ANOVA\s with age as between factor and numerosity as within factor were conducted per configuration to further investigate the interaction between configuration, age, and numerosity. The interaction between age and numerosity was significant in random configurations, $F(5, 60) = 3.63$, $p = .006$, $\eta_p^2 = .23$, but not in dice, $F(5, 100) = 1.67$, $p = .149$, and line configurations, $F(5, 55) = 1.31$, $p = .131$. Figure 2 shows that 4-year-olds had higher error rates than 5-year-olds in random configurations, on problems with five and six elements.

\(^1\) Chi-square tests were performed as well because distributions of error rates were possibly not normal. A chi-square test, testing the dependence of configuration and error rates was performed by number, by duration. Alpha level was divided by the number of chi-square tests per duration condition (6) and adjusted to .008. In the limited time condition, configuration and error rates were independent for numerosities one, two, three, and four, $\chi^2(2) = 0.020$, $p = .990$; $\chi^2(6) = 5.22$, $p = .516$; $\chi^2(6) = 11.15$, $p = .004$; $\chi^2(8) = 17.86$, $p = .022$, for 1–4 elements, respectively. Configuration and error rates were clearly dependent for numerosity five, $\chi^2(8) = 23.17$, $p = .003$. Cramer’s $V = .49$, and six, $\chi^2(8) = 20.80$, $p = .008$, Cramer’s $V = .46$. Error rates were lower than expected for dice configurations and higher than expected for random and line configurations. In the unlimited time condition, configuration and error rates were independent, $\chi^2(4) = 2.56$, $p = .634$; $\chi^2(4) = 3.92$, $p = .417$; $\chi^2(6) = 8.37$, $p = .212$; $\chi^2(8) = 12.05$, $p = .149$; $\chi^2(8) = 5.66$, $p = .685$, for 2–6 elements, respectively. No errors were made on problems with one element, in the unlimited time condition. Hence, results of the chi-square tests replicated results of the ANOVA\s.

\(^2\) Chi-square tests, testing the dependence of error rates and duration, were performed by number, for random configurations. Alpha level was adjusted to .008. Duration and error rates were independent for one, two, and three elements, $\chi^2(4) = 2.15$, $p = .71$; $\chi^2(4) = 1.04$, $p = .90$; $\chi^2(4) = 6.09$, $p = .19$. Duration and error rate were dependent for four, $\chi^2(4) = 15.43$, $p < .01$, Cramer’s $V = .74$, five, $\chi^2(4) = 13.16$, $p = .01$, Cramer’s $V = .69$, and six elements, $\chi^2(4) = 18.24$, $p < .01$, Cramer’s $V = .81$. Lower error rates on the unlimited, compared with the limited time task, caused the dependence. Results of the chi-square tests replicated results of the ANOVA\s.
Variations in configuration did not affect error rates when children enumerated 1–4 elements. Error rates were low for all configurations, in both the limited and the unlimited time condition. From five elements, however, dice presentations of elements were significantly easier compared with random or line presentations but only in the limited time condition. Time limits affected error rates from four elements. Hence, task manipulations affected performance in the counting range, but not the subitizing range. Children probably estimated the number of elements in the counting range, when in the limited time condition, because counting was impossible. Four-year-olds demonstrated a more steep increase in error rate than 5-year-olds from five elements, when enumerating problems with a random display. In conclusion, these results support the hypothesis that distinct processes are used for enumeration of elements in the two ranges (Trick, 2008), but not when elements are presented in dice patterns.

STUDY 2

Method

Participants

Data were collected between September 2010 and March 2013 with the ‘enumeration game’, part of project Math Garden (Klinkenberg, Straatemeier, & van der Maas, 2011). Math Garden is a Web-based computer-adaptive practice and monitoring system, used at
school and at home by over 60,000 children. Children practice math skills by playing math games, linked to plants in a personal garden. Playing a game makes the plant grow. Here, we describe only those aspects that are essential to understand the data analyses below. Participating schools gave permission to use data from their students for research purposes and accepted responsibility to inform parents accordingly. Parents of private individuals (a minority of the sample) electronically signed for permission for use of their data in scientific research. Four-, 5-, and 6-year-olds were selected. Table 1 shows distribution of age and gender of participants who attempted at least one of the problems that were analysed. Both problem difficulties and RTs per problem were subject to analysis. Estimation of problem difficulties is explained in the next section.

**Material and procedure**

Fifteen enumeration problems were presented sequentially in a session, see Figure 3 for an example problem. For each problem, two consecutive screens appeared. The first

Table 1. Distribution of age groups, gender, and average number of attempted problems in selected set in enumeration game of Math Garden

<table>
<thead>
<tr>
<th>Age</th>
<th>N (% of males)</th>
<th>Average number of attempted problems in selected set (standard deviations in brackets)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>1,285 (52.8)</td>
<td>93.1 (106.9)</td>
</tr>
<tr>
<td>5</td>
<td>3,364 (49.6)</td>
<td>69.2 (99.2)</td>
</tr>
<tr>
<td>6</td>
<td>9,778 (50.7)</td>
<td>26.7 (54.1)</td>
</tr>
<tr>
<td>Total</td>
<td>14,427 (50.6)</td>
<td>52.3 (91.9)</td>
</tr>
</tbody>
</table>

*Note.* There were 12,302 unique players with at least one response on the selected problems. A total of 2,125 children played at least one particular problem in two different age groups.

Figure 3. Study 2: Example of an enumeration problem in Math Garden. The example shows a random display of five elements.
screen showed the elements to be enumerated, a clickable keyboard with numbers 1–10, a question mark in case a participant did not know the answer, a coin bag, a row of 20 coins, and a green bar indicating game progress. Presentation duration was user-terminated, with a maximum of 20 s. A coin disappeared with each expiring second. Users clicked a response, which started the presentation of the second screen (1,000 ms in case of a correct response; 3,000 ms in case of an error), showing the correct response in green font and, if applicable, the false response in red font. After a correct (incorrect) response, the total in the coin bag was increased (decreased) by the number of remaining coins, with the notation that the total could not become negative. Hence, children were rewarded for fast accurate answers but penalized for fast inaccurate answers (Klinkenberg et al., 2011; Maris & Van der Maas, 2012). Coins could be spent on prizes in a virtual trophy cabinet. If the question mark was clicked, no coins were won or lost and the correct response turned green.

Selection of each enumeration problem resulted from a match between problem difficulties and participants’ enumeration skills. Both were estimated simultaneously using a computer-adaptive method (Klinkenberg et al., 2011), based on the Elo algorithm (Elo, 1978). After a correct response, the estimated problem difficulty lowered (dependent on RT and estimated participant’s skills) and the estimated skill increased (dependent on RT and estimated problem difficulty). The reverse happened after an error. Selection of the next problem depended on adjusted estimations of skill and problem difficulties. Over time, problem difficulties converged to a stable level. The next problem could be of different numerosity and/or different configuration but was chosen such that the average expected probability of a correct answer equaled .75. Hence, order of problems differed across participants.

Here, we focused on problems with 1–6 elements, in random, dice, or line configuration. Because of Math Gardens’ adaptive nature, we did not collect data of children who were able to count larger numbers quickly and accurately because they were presented with more difficult items. Table 2 shows the number of available problems, by numerosity.

\[^3\text{Problem difficulties and participants’ skills are estimated simultaneously and updated continuously after a first choice of starting values. Starting values are based on problem size for problem difficulties and on age for participants’ skills. Updating of a problem’s difficulty and a participant’s skills happens after the participant has solved a problem, according to Equation 1:}\]

\[\begin{align*}
\Theta_j &= \Theta_j + K_j (S_j - E(S_j)) \\
\beta_i &= \beta_i + K_i (E(S_i) - S_i)
\end{align*}\]

where \(\Theta_j\) is the skill estimate of participant \(j\); \(\beta_i\) is the difficulty estimate of problem \(j\); \(S_j\) and \(E(S_j)\) are the score and expected probability of succeeding for person \(j\) on problem \(i\). \(K\) is a function of the problem difficulty uncertainty \(\mathcal{U}\) of the participant and the problem (Equation 2):

\[K_j = K(1 + K_+ \mathcal{U}_j - K_- \mathcal{U}_j)\]

\[K_i = K(1 + K_+ \mathcal{U}_i - K_- \mathcal{U}_i)\]

where \(K = 0.0075\) is the default value when there is no uncertainty; \(K_+ = 4\) and \(K_- = 0.5\) are the weights for the estimate uncertainty for participant \(j\) and problem \(i\). \(\mathcal{U}\) is uncertainty, which depends on both recency (the more recent, the lower the uncertainty) and frequency (the higher the frequency, the lower the uncertainty) of playing. Klinkenberg et al. (2011) assume that uncertainty reduces to 0 after 40 administrations, but increases to the maximum of 1 after 30 days of not playing (Equation 3).

\[\mathcal{U} = \mathcal{U} \frac{1}{40} + \frac{1}{30} D\]

Where \(D\) refers to the number of days without playing.

A more elaborate description of the procedure of estimating problem difficulties and participants’ skills is given in Maris and Van der Maas (2012) and Klinkenberg et al. (2011).
Math Garden’s computer-adaptive method is based on estimation of problem
difficulties and individual’s skills on one and the same scale, across age groups. However,
for Study 2, problem difficulties were recalculated for 4-, 5-, and 6-year-olds separately,
using children’s logged records of problems attempted, their answer, and RT on each
problem. The estimation procedure was rerun by age group and resulted in a difficulty and
an average RT for each problem, for each age group. The Euclidean distance, which is the
straight-line distance between two elements, based on the grid used in Math Garden, was
averaged across all elements, for each display and included as a covariate in subsequent
analyses.

Results

Problem difficulties
Although the number of problems may appear low, we contend that problem
difficulties reflect participants’ varying performance in enumeration, dependent on
numerosity, and configuration because problem difficulties were based on responses
from a large sample (see Table 1), had converged to stable values, and had small
standard errors. Also, split-half reliability of problem difficulties, based on a split of the
problems into two groups, with numerosity and configuration equally divided, was
high, $r = .98$, $p < .001$.

Left panels in Figure 4 graph problem difficulties against numerosity for random, dice,
and line configurations, by age group. Problems with one element are graphed in Figure 4
but excluded from analyses because they could not be assigned exclusively to one
configuration. A univariate ANOVA with factors configuration (random, dice, and linear),
age group (4–6 years), and numerosity (2–6), covariate average Euclidean distance, and
dependent variable problem difficulties was performed. Average Euclidean distance did
not affect problem difficulty, $F(1, 197) = 0.28$, $p = .60$. The main effect of configuration
was significant, $F(2, 197) = 72.11$, $p < .001$, $\eta^2_p = .50$. Post-hoc analyses indicated that
problem difficulties were significantly lower for dice compared with random and line
configurations ($p < .001$). The main effect of age was not significant, $F(2, 197) = 0.11$, $p = .89$. This is a consequence of the recalculation method in which the average problem
difficulty was set to 0 for each age group. The main effect of numerosity was significant,
$F(4, 197) = 248.73$, $p < .001$, $\eta^2_p = .25$. Post-hoc analyses indicated that each additional

Table 2. Number of problems in selected set in enumeration game of Math Garden, by numerosity, by
display

<table>
<thead>
<tr>
<th>Numerosity</th>
<th>Display of elements</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Random</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
</tr>
</tbody>
</table>

Note. aProblems with one element could not be assigned exclusively to one of the configurations.
element increased problem difficulty significantly ($p < .05$ for comparisons between displays of two subsequent numbers).

Main effects were qualified by the interaction between configuration and numerosity, $F(8, 197) = 7.66, p < .001, \eta^2_p = .12$. Post-hoc analyses indicated that the configuration

Figure 4. Study 2: Problem difficulties and average response times (RTs) in Math Garden by numerosity, configuration, and age group. Left panels show problem difficulties, and right panels show average RTs.
effect was not significant for two and three elements, $p > .05$. For four, five, and six elements, problems with dice displays were easier than problems with random and line displays, $p < .001$, whereas problem difficulties of the latter did not differ from each other for all numerosities, $p > .05$. The interaction between age group and numerosity was significant as well, $F(8, 197) = 7.76, p < .001, \eta^2_p = .13$. In 4-year-olds, problems with two elements were significantly easier than problems with three elements, which were significantly easier than problems with four elements, $p < .001$. Difficulties of problems with four, five, and six elements were equally difficult in 4-year-olds, $p > .05$. In 5- and 6-year-olds, problems with two and three elements were equally difficult, $p > .05$, whereas problem difficulty increased significantly from 3 to 4 and from 4 to 5 elements. The interaction between configuration and age group was not significant, $F(4, 197) = 2.11, p = .081$. The three-way interaction was also not significant, $F(16, 197) = 1.61, p = .069$.

Summarized, dice displays, as compared to random and line displays, only lowered problem difficulties in the counting range. Moreover, 5- and 6-year-olds, but not 4-year-olds, showed the typical pattern of comparable difficulties for problems with two and three elements and increasing difficulty with each additional element, up to six.

Problem difficulties are the result of a combination of RTs and accuracy. To align with previous studies, an RT analysis is reported next. Error rates are not reported because the adaptive algorithm was designed to keep error rate at a constant value of .25 for as much as possible, for all individuals.

**Response times**

The right panels of Figure 4 show RTs averaged by configuration, numerosity, and age group. A univariate ANOVA with factors configuration, age, and numerosity, and covariate average Euclidean distance was performed with RTs as dependent variable. Average Euclidean distance did not significantly influence RTs, $F(1, 197) = 0.02, p = .892$. The main effect of configuration was significant, $F(2, 197) = 195.63, p < .001, \eta^2_p = .50$: RTs were longer for problems with random as compared to line displays, which were longer than RTs of problems with dice displays, all $p < .001$. The significant main effect of age, $F(2, 197) = 576.36, p < .001, \eta^2_p = .50$ indicated that RTs decreased with age, $p < .001$. Finally, the significant main effect of numerosity, $F(4, 197) = 764.13, p < .001, \eta^2_p = .25$, showed that RT increased with each additional element, all $p < .001$.

Main effects were qualified by significant interactions between configuration and numerosity, $F(8, 197) = 30.21, p < .001, \eta^2_p = .13$, and between age and numerosity, $F(8, 197) = 3.50, p < .001, \eta^2_p = .13$. The interaction between age and configuration was not significant, $F(4, 197) = 2.18, p = .072$, just as the three-way interaction, $F(16, 197) = 0.67, p = .826$. Post-hoc analyses indicated that the configuration effect was not significant for two and three elements, $p \geq .968$, but RTs were lower for problems with dice as compared to random and line displays for four, five, and six elements, $p < .001$. RTs for line and random display problems did not differ significantly from each other, $p \geq .056$. Post-hoc analyses also indicated that in 4- and 5-year-olds, RTs for enumerating two and three elements were equal, $p \geq .708$, but RT increased with each additional element, $p < .001$. In 6-year-olds, however, RT increased with each additional element, $p \leq .002$. 


Conclusion
In general, problem difficulties and RTs increased with increasing number of elements, decreased with increasing age, and were lower for problems with dice as opposed to random or line displays, but effects of configuration, age, and numerosity interacted. The configuration effect only occurred when the number of elements was four or higher. In 4-year-olds, problem difficulty increased with each additional element, whereas RTs were equal in the subitizing range, but increased with each additional element in the counting range. In 5-year-olds, both problem difficulties and RTs were equal in the subitizing range and increased with each additional element in the counting range. In 6-year-olds, problem difficulties were equal in the subitizing range and increased with each additional element in both ranges. Taken together, the results of the analyses of problem difficulties and RTs converged and suggest that manipulating the configuration of elements affected performance for problems in the counting range only and that the maximum number of elements that children could subitize was three, in all three age groups.

Discussion
In a long-standing discussion, it is debated whether humans use a single or two different processes to precisely enumerate small versus large numbers, referred to as subitizing and counting (Mandler & Shebo, 1982). The ranges in which the processes are used are referred to as the subitizing and the counting range. Here, we follow the argument that the claim of the existence of two distinct processes is supported, when task manipulations have different effects in the subitizing compared with the counting range (Trick, 2008).

In Study 1, a sample of Dutch 4- and 5-year-olds enumerated visually presented dots. Dots were arranged randomly, in a line or in a familiar pattern (dice). Presentation duration was limited or unlimited. Study 2 included an analysis of problem difficulties and RTs, obtained with a computer-adaptive math program (Klinkenberg et al., 2011). Configuration was manipulated as in Study 1. In both studies, task manipulations did not affect performance in the subitizing range, but did affect performance in the counting range. This consistent finding in these markedly different studies supports the claim that subitizing and counting are distinct processes (Mandler & Shebo, 1982; Schleifer & Landler, 2011; Trick, 2008).

The findings on development in subitizing range diverged. In Study 1, 5-year-olds, but not 4-year-olds, show a clear difference between performance in the subitizing as opposed to the counting range. In Study 2, problem difficulties and RTs complemented each other and showed that 4-, 5-, and 6-year-olds were able to enumerate up to three elements fast and accurately and probably resorted to counting or estimation for larger numerosities.

Results on enumeration of elements in line configurations show that performance on problems with line and random configurations was comparable. If anything, problems with line configurations were easier than problems with random configurations. Line configurations might facilitate counting, as it is easy to move from one to the next element and to remember which elements are already counted. Balakrishnan and Ashby (1992) solely used linear configurations and found that performance continuously decreased with increasing numerosity, also in the subitizing range. Differences between their study and the present studies demonstrate the necessity of varying the configuration of elements.
Both studies show that arranging elements in dice configurations facilitates enumeration of large numbers of elements compared with random configurations. Performance in the subitizing and counting range was similar when elements were presented in dice configurations (see Mandler & Shebo, 1982 who made a similar observation for adults). These results suggest that pattern recognition can help the enumeration of large numbers. Children with developmental dyscalculia show a deficit in both subitizing and the fast enumeration of elements in the counting range, when presented in familiar patterns. Difficulties in pattern recognition may relate to these deficits (Ashkenazi, Mark-Zigdon, & Henik, 2013). Visuo-spatial working memory (VWM) is a prerequisite for pattern recognition (Ashkenazi et al., 2013). Hence, the hypothesis that subitizing is based on pattern recognition matches the observation that VWM capacity correlates significantly with subitizing capacity (Piazza, Fumarola, Chinello, & Melcher, 2011). Although subitizing may also be the result of the application of a limited number of spatial indexes (FINSTs; Trick & Pylyshyn, 1994), Vetter, Butterworth, and Bahrami (2008) show that it cannot be a pre-attentive process.

In Study 1, performance in the counting range increased when given unlimited observation time as compared to limited time. However, performance for dice patterns was already at ceiling in the limited time condition. Note that the number of possible patterns for numerosities in the subitizing range is limited and that many possible random patterns map the familiar dice patterns. Hence, associating patterns with number words is relatively easy in the subitizing range. The number of possible patterns grows exponentially in the counting range (Benoit et al., 2004), complicating the association between patterns and number words. As a consequence, children in the studied age range may have already learned these associations in the subitizing range, but not yet in the counting range. This sketch of development implies that counting is a prerequisite for subitizing (Gelman & Gallistel, 1978).

Mandler and Shebo (1982) and Ashkenazi et al. (2013) already proposed that subitizing is based on acquisition of associations between patterns and number words. The finding that subitizing improves with age (Maylor et al., 2011; Reeve et al., 2012; Starkey & Cooper, 1995; Trick, Enns, & Brodeur, 1996) supports this hypothesis. However, the hypothesis conflicts with results from infant studies (Carey, 2004; Desoete, Ceulemans, Roeyers, & Huylebroeck, 2009; Feigenson, Dehaene, & Spelke, 2004) and studies on number discrimination among people in remote cultures (Dehaene, 1987). Studies including a broad age range, applying a shared paradigm across age groups, may contribute to the discussion on the origins of subitizing.

The current studies are not without limitations. In Study 1, only four repetitions of each number were used in each condition, sample size was small, and distance between elements did not vary randomly across numerosities. Fortunately, Study 2 showed that distance between elements did not influence problem difficulties. In Study 2, administration was unsupervised, and order of presentation of numerosities and displays was uncontrolled. However, the large sample size probably averages out effects of environment and order.

Summarized, the results suggest that children use distinct processes for the enumeration of small and large sets of elements when presented in random patterns. The results concerning enumeration of elements in dice patterns suggest that performance of preschoolers can be explained by three processes: Counting (for large numbers, when given sufficient time), estimation (for large numbers, when time is limited), and subitizing. The latter might be based on pattern recognition as performance for dice patterns is comparable for small and large numbers. If subitizing indeed involves
pattern recognition, extension of subitizing to larger numbers may be possible (as the recognition that two dice patterns of six represent 12 elements), comparable to the recognition of complex chess patterns by advanced chess players (De Groot, 1978; Simon & Chase, 1973). Training of pattern recognition (see Fischer, Königeter, & Hartnegg, 2008; Wilson, Revkin, Cohen, Cohen, & Dehaene, 2006) and presenting elements in fixed patterns may ease number recognition and encourage insight into simple addition and subtraction. After all, enumeration is an important requisite for later math skills (Jordan et al., 2007; Krosbergen et al., 2009).

Acknowledgements

The authors thank Wesley Bonda, Barbara de Ridder, Raïsa Koolbergen, and Cindel van der Valk for their assistance in collecting the data.

References


Received 19 November 2012; revised version received 12 April 2013