

Range dependent processing of visual numerosity: similarities across vision and haptics

Myrthe A. Plaisier · Wouter M. Bergmann Tiest ·
Astrid M. L. Kappers

Received: 17 February 2010 / Accepted: 27 May 2010 / Published online: 12 June 2010
© The Author(s) 2010. This article is published with open access at Springerlink.com

Abstract ‘Subitizing’ refers to fast and accurate judgement of small numerosities, whereas for larger numerosities either counting or estimation are used. Counting is slow and precise, whereas estimation is fast but imprecise. In this study consisting of five experiments we investigated if and how the numerosity judgement process is affected by the relative spacing between the presented numerosities. To this end we let subjects judge the number of dots presented on a screen and recorded their response times. Our results show that subjects switch from counting to estimation if the relative differences between subsequent numerosities are large (a factor of 2), but that numerosity judgement in the subitizing range was still faster. We also show this fast performance for small numerosities only occurred when numerosity information is present. This indicates this is typical for number processing and not magnitude estimation in general. Furthermore, comparison with a previous haptic study suggests similar processing in numerosity judgement through haptics and vision.

Keywords Subitizing · Visual perception · Numerosity judgement

Introduction

In visual numerosity judgement, three different processes can be identified. For exact numerosity judgement it has been suggested that there are different processes for judging small and large numerosities. Small numerosities (≤ 4) are judged fast and error-free through a process that has been labeled ‘subitizing’ (e.g. Kaufman et al. 1949; Atkinson et al. 1976; Mandler and Shebo 1982; Trick and Pylyshyn 1993). The slope of the response times as a function of the number of items in this regime is generally found to be 40–100 ms/item (e.g. Akin and Chase 1978; Oyama et al. 1981; Trick and Pylyshyn 1993; Trick 2008). For larger numerosities (> 4) the slower and more error-prone process of ‘counting’ is used and response times and error rates increase with the number of items. The slopes of the response times are usually 200–400 ms/item in this regime. Note that although counting is thought to be more error-prone than subitizing, it can be very precise provided that there are no restrictions on time. In addition to precise numerosity, humans adults, but also children, can judge approximate numbers (e.g. Beran et al. 2006; Whalen et al. 1999; Dehaene et al. 1998). Judging approximate numerosity without counting is an ability that has also been shown to exist in animals such as monkeys, dogs, pigeons, parrots and fish (e.g. Boysen 1997; Roberts et al. 2002; West and Young 2002; Pepperberg 2006; Agrillo et al. 2007). In primates it has been shown that there are neurons tuned for specific numerosities (Nieder et al. 2002). This suggests number representation is innate. This fast process for judgement of approximate numerosity will be referred to as ‘estimation’. Numerosity judgements through estimation become less precise for larger numerosities and obey Weber’s law stating that precision is a constant fraction of the magnitude. Therefore, discriminability of

M. A. Plaisier · W. M. B. Tiest · A. M. L. Kappers
Helmholtz Institute, Universiteit Utrecht,
Padualaan 8, 3584 CH Utrecht, The Netherlands

Present Address:

M. A. Plaisier (✉)
Faculty of Human Movement Sciences, VU University,
Van der Boechorststraat 9, 1081 BT Amsterdam,
The Netherlands
e-mail: m.plaisier@fbw.vu.nl

two numerosities is defined by their ratio (Izard and Dehaene 2008; Gallistel and Gelman 1992, 2000). On a neurological level it has also been shown that number is encoded following Weber's law. In monkeys it has been reported that populations of numerosity-selective neurons encode each number only approximately with an imprecision that increases with the number (Nieder and Miller 2003). To make a numerosity judgement, this internal continuous representation of magnitude still has to be mapped onto an arabic numeral or number word (Whalen et al. 1999; Moyer and Landauer 1967). This mapping has some variability as magnitude representations are retrieved from memory. Recently, it has been shown that this mapping can be re-calibrated by providing feedback after each numerosity judgement (Izard and Dehaene 2008).

The question of what kind of a process subitizing actually is has yet to be answered. It has been suggested that it is not a separate process at all. Balakrishnan and Ashby (1992) have suggested that there is no evidence for the existence of a subitizing regime. Others have argued that subitizing is caused by large relative differences between small numerosities (Van Oeffelen and Vos 1982). For instance, the relative difference between 2 and 3 is much larger than between 6 and 7. It has been shown that there is a 25% Weber fraction for the discrimination of large numerosities (8–64 items) (Ross 2003). This would explain a transition to counting above four items, because then the relative difference between subsequent numerosities becomes smaller than the discrimination threshold. Recently, it has been shown that the hypothesis that subitizing is very accurate estimation does not hold (Revkin et al. 2008). In that study, the authors compared judgement of 1, 2, 3, 4, 5, 6, 7 or 8 items to judgement of 10, 20, 30, 40, 50, 60, 70 or 80 items. Note that the relative differences between subsequent numerosities were the same for both numerosity ranges. By limiting the response time, subjects were prevented from counting the items. For the first range they found that judgement of 1 to 4 items was faster and more accurate than for the larger numerosities. In contrast, for the second range there was no clear advantage for numerosities 10 to 40 compared to 50 to 80. This suggests that subitizing is not a Weberian estimation process.

Cordes et al. (2001), however, did not find such a discrepancy between subitizing and counting range in a study where subjects were shown a numeral and had to make the corresponding number of key presses with verbal and non-verbal counting. In the verbal counting condition, subjects counted the number of key presses out loud, while in the non-verbal condition they had to say "the" with every key press. The coefficient of variation (ratio between the mean response and the standard deviation) was constant over the whole range in both conditions indicating that there was no special performance for small numbers. This suggests that

small numbers are represented in the same way as larger numbers, which contrasts the study by Revkin et al. (2008). This could be due to the fact that in the Revkin et al. study, numbers were represented by a collection of dots, while in the Cordes et al. study numerals were used and the according number of key presses had to be made. Subitizing may only be relevant for processing sets of spatially distributed items. Logie et al. (1987) reported interference of articulatory suppression (saying "the") for judgement of items distributed in time (flashes), but not for judgement of spatially distributed items. Furthermore, interference of finger tapping was smaller than for articulatory suppression for temporal numerosity judgement. A systematic study of spatial judgement of number showed that finger tapping interferes more than articulatory suppression (Trick 2005). This suggests that not all types of numerosity information are processed in the same way. For sets of spatially distributed dots it has been shown that sets of dots from the subitizing range are rated as more dissimilar than larger sets of dots (Logan and Zbrodoff 2003). These findings indicate that when dots scattered over a display are shown, for some reason numerosities from the subitizing range are recognized faster and more accurately than larger numerosities.

Although it is not clear what causes the fast and accurate judgement of small sets of dots, it has been shown that subitizing is not limited to visual numerosity judgement. Subitizing has been shown to occur for up to two items in audition (Ten Hoopen and Vos 1979; Camos and Tillmann 2008). Note however, that in this case items are often presented sequentially instead of simultaneously. More recently, subitizing has also been shown to exist in haptic numerosity judgement for both 'passive touch' (i.e. touch without active exploration) (Riggs et al. 2006), as well as 'active touch' (Plaisier et al. 2009). In this last study, we have addressed the role of the relative differences between subsequent numerosities in the numerosity range. Subjects had to grasp and judge 1, 2, 4, 8 or 16 spheres. Note that there was always a factor of 2 between subsequent numerosities. In this case, we found that numerosity judgement was fast for all numerosities, but judgement of small numerosities (≤ 4) was even faster than for larger numerosities. We compared response times and error rates from this task to a different task in which subjects had to label single spheres varying in size. In this case no clear advantage for small sphere sizes was found. The response times from this second task could be described using a model based on Fechner's law for discriminability. This showed that discriminability followed the psychophysical power law over the whole range of sphere sizes. However, this model could not describe the pattern in the response times from the first task in which numerosity was varied. This suggests that although the relative differences

between subsequent numerosities were constant over the whole range, small numbers were recognized faster and more accurately than large numbers. Furthermore, this fast recognition was not mediated through the use of volume or mass cues.

In short, our haptic study showed that numerosity judgement without counting was faster for numbers from the subitizing range than outside this range, even when the relative spacing between numerosities was a factor of two over the whole range and feedback was provided so subjects could re-calibrate their number mapping. These results are in agreement with the study of Revkin et al, suggesting that subitizing is not the same process as estimation of large numbers. Based on this hypothesis, the results from our haptic study should be reproducible in the visual domain. Note that this approach is different from the one Revkin et al. (2008) used. In their study relative differences between subsequent numerosities varied over the stimulus range and subjects were forced to use estimation by limiting response times. Our approach is to make the relative differences between subsequent numerosities constant and larger than the discrimination threshold over the whole range. Therefore, subjects would be able to accurately judge the numerosity without counting over the whole range and will use estimation without being forced to do so. If our haptic data are reproducible in the visual domain, this is further support for the idea that numerosities from the subitizing range are recognized faster than outside this range and that this is not due to the mapping of numbers being increasingly less precise for larger numerosities. Moreover, it would argue for a shared representation of number between the visual and the haptic modalities. This has interesting consequences for the possible mechanisms underlying fast recognition of numbers in the subitizing regime as typical visual explanations, such as pattern recognition would in that case be very unlikely.

In Experiment 1, a ‘classic’ numerosity judgement task was performed in which we reproduce the well-known upward bend in the response times at about four items. To investigate what the effect was of decreasing relative differences for larger numerosities in Experiment 1, Experiment 2 was performed. Here, we presented subjects with numerosities that were chosen such that the relative difference between subsequent numerosities was constant over the whole range (1, 2, 4, 8, 16 or 32 items). Note that in this case relative differences between subsequent numerosities were larger than the discrimination threshold of 25% for judging number without counting. If subitizing were accurate estimation made possible because relative differences are above the discrimination threshold, we would not expect faster performance for small numerosities than for large numerosities. In the next experiment we investigated how response times scale with magnitude in

the absence of numerosity information. To this end, numerosity information was removed in Experiment 3 and subjects had to name dots with varying sizes. In this case one could expect response times to be constant over the whole range. However, in our haptic study we found end effects at both ends of the range. We also expect to find such effects here and used a model from our haptic study to account for these effects.

The first three experiments were a transference of our haptic experiments to the visual modality, but in Experiments 4 and 5 we go beyond that study. It has been suggested that humans have a shared representation of number and physical magnitude (Walsh 2003). If this is true for numbers outside the subitizing range we expect performance similar to that for dot size recognition. Therefore, in Experiment 4 we investigated whether response times for recognition of numbers outside the subitizing range (8, 16, 32, 64 or 128 items) follow the same pattern as those for dot size recognition. This would indicate that mapping of physical magnitude is shared with mapping of numerosities outside the subitizing range. If discriminability for large numbers follows the power law we do not expect a special regime for the smallest numerosities in the range in this case. Finally, in Experiment 5 numerosities from the subitizing regime were added to the numerosity range from Experiment 4 and we investigated how this affected recognition of the larger numerosities in the range. If discriminability of small numbers is indeed much better than that of large numbers, we expect that adding numbers from the subitizing range will not affect recognizability of the larger numerosities.

General method

Participants

Ten paid subjects (age 21 ± 3 years) participated in Experiments 1, 2, and 3. Five of them were female. They performed the three experiments in counterbalanced order. Ten other paid subjects (age 21 ± 2 years) participated in Experiments 4 and 5. Two of them were male. They performed the two experiments in counterbalanced order. All participants had normal or corrected to normal vision. They were treated in accordance with the local guidelines and gave their informed consent.

Set-up and procedure

Stimuli were presented on a 20 inch LCD monitor (Apple Cinema) with a 1050×1680 pixels resolution. A mask was placed over the monitor, leaving a circular display area with a diameter of 25 cm. Varying numbers of black dots

were presented on a white background. The circular area over which the dots were randomly distributed could be varied and will be referred to as the presentation area. The display was controlled using a LabVIEW program running under Mac OS. Time measurement was started when the dots appeared on the screen and was terminated when a vocal response was registered using a microphone. Through this system, response times were recorded with an accuracy of up to 3 ms.

Subjects were seated in a dark room at a distance of 57 cm from the monitor with their chin in a chin rest. At this distance an image of 1 cm on the monitor corresponded to 1° visual angle. First a fixation cross appeared in the centre of the display. After 1 s the cross disappeared and the stimulus was presented. The stimulus remained visible until a response was registered after which the stimulus disappeared. Subjects were instructed to respond as fast as possible either the number of dots (Experiments 1, 2, 4 and 5) or the dot size (Experiment 3) that was presented. It was also emphasized that it was important that the answer was correct. After each trial the experimenter entered the response into the computer and feedback on whether the answer was correct was shown on the screen for 1 s in all experiments. If the answer was incorrect, also the correct response was shown. Each experiment was preceded by a training session before the experiment was started. Subjects performed at least 20 training trials and training trials were continued until 10 in a row were answered correctly.

Analysis

Because subjects were instructed to respond correctly and therefore minimize their errors, the error rates should be low in all experiments. Also in the subitizing regime the error rate should be roughly zero. Therefore, error rates are shown as an indication that subjects could perform the task correctly and the response times were used for further analysis. Response times of incorrectly answered trials were excluded from the analysis. Also, response times that deviated more than 3 SD from the mean were discarded as outliers. When sphericity was violated in the statistical analysis, Greenhouse-Geisser corrected values are reported. When the analysis involved regression, we report the results from the regression to the response times averaged over subjects. We also report the mean parameter values determined through regression of the model to the single subjects' data. Note that this does not necessarily yield the same outcome. Regression to the data averaged over subjects is more accurate, but it is also important to show that the same trend is present in the data for each subject individually. Therefore, the results from both procedures are reported. In all regression procedures the response

times were weighted according to their inverse squared standard deviations.

Experiment 1

The purpose of this experiment was to validate our experimental paradigm (e.g. Mandler and Shebo 1982; Trick and Pylyshyn 1993). In order to do so, we reproduce the classical two regimes in visual numerosity judgement for small and larger numerosities. The slope of the response times as a function of the number of items and the transition point from subitizing to counting may depend on the stimulus and varies among subjects. This experiment was performed to determine these values for the specific stimulus used in this particular experimental design and this pool of subjects.

Method

Stimuli

In this experiment 1, 2, 3, 4, 5, 6, 7, 8 or 9 black dots were presented on a white background. The dots had a diameter of 0.5° and the presentation area had a diameter of 20°. The dots were placed such that their edges were at least 0.8° apart and 0.8° from the edge of the presentation area. Each numerosity was presented 16 times.

Analysis

To accurately determine the values of the slopes in the subitizing and counting regimes without making assumptions about the location of the transition point between the regimes, regression of a bilinear model was used. The bilinear function is given by:

$$T(N) = (r_1N + c_1)H\left(\frac{c_2 - c_1}{r_1 - r_2} - N\right) + (r_2N + c_2)H\left(N - \frac{c_2 - c_1}{r_1 - r_2}\right). \quad (1)$$

where N is the number of items, $H(N)$ is the Heaviside step function and r_1 and r_2 are the slopes, while c_1 and c_2 represent constant offsets. Note that through this analysis the location of the transition point follows from the intersection of the two linear parts and is given by:

$$N_t = \left(\frac{c_2 - c_1}{r_1 - r_2}\right). \quad (2)$$

The last data point at nine items was not included in the regression analysis, because of possible end-effects. Subjects usually learn what the maximum numerosity is during the experiments, so after counting the first 8 items they

already know that the answer is 9. This reduces response times and this might lead to deviations from linearity for the response times of the largest numerosity in the range. Excluding the largest numerosity is commonly done in numerosity judgement studies (e.g. Trick and Pylyshyn 1993; Watson et al. 2007; Trick 2008).

Results

The response times and error rates averaged over subjects are shown in Fig. 1. It can be seen that numerosity judgement was error-free for up to four items. Repeated measures ANOVA on the response times with numerosity as within subjects factor, showed a significant main effect ($F(1.8, 16) = 148, P < 0.001$). Trend analysis showed that there was a significant deviation from linearity ($F(1, 9) > 23, P < 0.001$). Regression of the bilinear function to the response times averaged over subjects and weighted according to their standard deviation, yielded a slope of 46 ms/item for the first part of the stimulus range and a slope of 270 ms/item for the second part of the range ($R^2 = 0.99$). The transition point was located at 3.7 items, so in between 3 and 4 items.

As was mentioned before, the transition point and also the response time slopes may vary among subjects. Therefore, the response times were also analyzed for each subject separately. The bilinear model was fitted to the single subjects' response times. The slopes and transition points from the individual subjects were then averaged. This yielded a slope of 35 ± 9 ms/item (SE) for the first regime and 272 ± 17 ms/item (SE) for the second regime. The transition point was located at 3.6 ± 0.3 (SE) items. For four subjects the transition point was in between 4 and 5 items, three subjects had the transition point in between 3 and 4 items and two of the subjects had the transition point

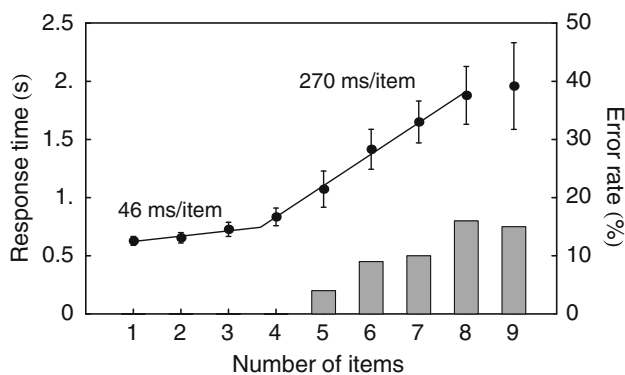


Fig. 1 Response times (dots) and error rates (bars) averaged over subjects from Experiment 1. The solid line represents the best fit of the bilinear function to the response times averaged over subjects. Slope values are indicated in the figure. The response time for nine items was not included in the regression analysis. Error bars indicate the standard deviation of the single subject means

between 2 and 3 items. The overall quality of the fits was good, $R^2 = 0.989 \pm 0.002$ (SE).

Discussion

The values of the subitizing and counting slopes found here are in agreement with the existing literature on numerosity judgement of 40–100 ms/item in the subitizing range and 200–400 ms/item in the counting range (e.g. Akin and Chase 1978; Oyama et al. 1981; Trick and Pylyshyn 1993; Trick 2008). Note that this does not necessarily mean that different processes are used for small and large numerosities. There could still be a single underlying process. Rather, these results show that our results are comparable to previous results.

It has been proposed that small numbers are somehow recognized fast and accurately, so there is no need to count them. A possible explanation for a transition from subitizing to counting is then that the relative differences between the subsequent numerosities become successively smaller. When the relative differences are large it may be easy to recognize a certain numerosity. If this were true, it is expected that also larger numerosities can be easily and accurately recognized if the presented numerosities are chosen such that the relative differences are large over the whole range. In that case, there should be no longer an advantage for small numerosities. This was investigated in Experiment 2.

Experiment 2

The purpose of this experiment was to investigate how response times were influenced by the relative differences between subsequent numerosities in the presented range. The numerosity range was chosen such that there was always a factor of two between subsequent numerosities, because this was the largest relative difference between subsequent numerosities in Experiment 1. We expect that subjects can recognize the different numerosities without counting and response times will be smaller than those found in the counting range in Experiment 1. If an advantage for small numerosities is found, this indicates that subitizing is not related to relative differences between the numerosities. To exclude the possibility that larger response times for larger numerosities were caused by a longer time needed to verbalize these numbers, a digit-naming experiment was carried out as a control.

Method

Subjects were shown 1, 2, 4, 8, 16 or 32 dots and they had to respond the number of dots. Subjects were explicitly told

which numbers could be presented before the experiment started. Dot diameter was the same as in Experiment 1 (0.5°) and the presentation area had a diameter of 20° . Also a control condition was performed in which subjects were shown digits forming the numbers: 1, 2, 4, 8, 16 or 32, in the centre of the screen and subjects had to respond by calling out the presented number. The height of a digit was 2° .

Results

Response times averaged over subjects and error rates for the different numerosities are shown in Fig. 2a. It can be seen that the responses were faster for small numerosities (<4), compared to larger numerosities. Repeated measures ANOVA on the response times showed that the effect for numerosity was significant ($F(1.2, 10.8) = 18.6$, $P < 0.001$). Trend analysis showed that there was a significant linear trend ($F(1, 9) = 23.3$, $P < 0.001$) and a significant cubic trend ($F(1, 9) = 24$, $P < 0.001$) in the response times. This indicates that there was an increase of the response times from small to larger numerosities, but there was also twice a change of direction of the trend. This

resulted in the S-like shape in the response times that can be seen in Fig. 2a. Regression of a linear function yielded a significant slope of 17 ms/item ($P = 0.03$, $R^2 = 0.7$).

The results for the digit-naming condition are shown in Fig. 2b. It can be seen that response times are relatively constant over the whole range and no errors were made. Repeated measures ANOVA on the response times showed that there was a significant effect of numerosity ($F(15, 45) = 17$, $P < 0.001$). However, the linear trend was not significant ($F(1, 9) = 1.4$, $P = 0.27$). Pair-wise comparisons showed that there were several significant differences between the different numbers. The largest average difference was 80 ms between numbers 4 and 8 ($P = 0.001$, Bonferroni corrected value).

Discussion

The control experiment showed that there was an effect of numerosity. But more importantly, there was no increase of the response times from small to large numbers. This shows that there was no difference in the time needed to verbalize small and large numbers. Therefore, this cannot explain the advantage in judgement of small numerosities.

In the main experiment response times were well below 1.5 s over the whole numerosity range, so subjects were clearly not counting the items. From Experiment 1 it can be seen that counting 8 items already takes 2 s. Therefore, we conclude that subjects could recognize the large numerosities (8, 16 and 32) without counting. The results show that when the relative differences between subsequent numerosities are large over the whole numerosity range, subjects can recognize all numerosities without counting. However, there was still an advantage for small numerosities. This shows that small numerosities were recognized faster than large numerosities for reasons other than the relative differences between subsequent numerosities. This is in agreement with what we found in our previous study on haptic numerosity judgement (Plaisier et al. 2009). To investigate what mediates this fast recognition of small numbers, Experiment 3 was carried out in which numerosity information was removed and only other magnitude information was present. It has been suggested that representation of number is shared with magnitude representation. If this fast performance for small numerosities is specific to number representation, we do not expect it to appear for the smallest stimuli in Experiment 3.

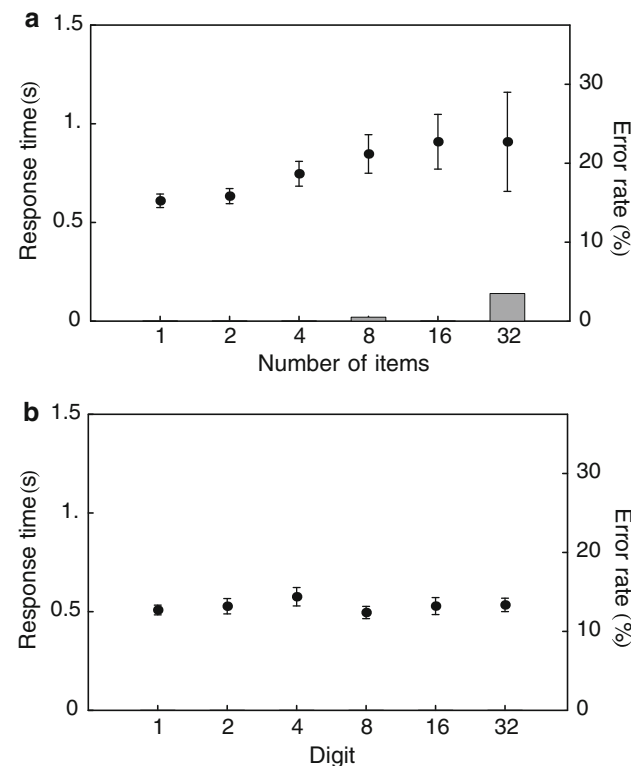


Fig. 2 **a** Response times (dots) and error rates (bars) averaged over subjects from Experiment 2. **b** Response times and error rates (these were zero for all numbers) averaged over subjects in the digit-naming condition. The error bars represent the standard deviation of the single subject means

Experiment 3

In this experiment subjects were shown a dot in the centre of the screen. The area of the dot always corresponded to the total area of one of the different numbers of dots from

Experiment 2. The dots were numbered accordingly and subjects had to respond the number that was associated with the dot size that was presented. Subjects could recognize the different dots by judging presentation area and luminance. These cues were also present in the stimuli of Experiment 2 and the only difference with respect to the stimuli of Experiment 2 is that the black pixels were all contained within a single disk around the centre instead of distributed over different disks. Consequently, if the fast recognition of small numerosities found in Experiment 2 was mediated by these cues, we expect that we will also find it in this experiment. If the special performance disappears we can conclude that the fast recognition of small numbers is related to black pixels being distributed in a certain way.

Method

Subjects were shown dots that had an area equivalent to the total area of the varying numbers of dots in Experiment 2. They had to respond with the corresponding label. For instance, when subjects saw the dot with area corresponding to the area of 4 dots in Experiment 2 (i.e. dot with diameter 1°), they had to respond by calling out 4. Consequently, the presented dots had a diameter of 0.5° , 0.7° , 1° , 1.4° , 2° or 2.8° . The subjects were shown the different dot sizes together with the labels before the training session was started. This mapping was not visible during the training session or experiment.

Results

Figure 3 shows response times and error rates averaged over subjects for the different dot sizes. Error rates were low (<20) over the whole stimulus range, indicating that subjects could perform the task correctly. Errors occur over the whole stimulus range in this case and not only for the largest

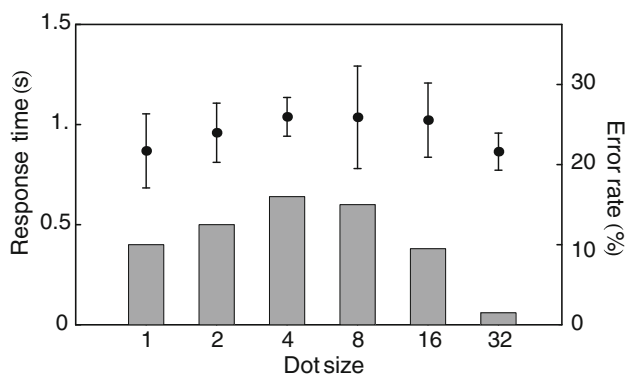


Fig. 3 Response times (dots) and error rates (bars) averaged over subjects for Experiment 3. The error bars represent the standard deviation of the single subject means

numerosities in the range like in Experiment 2. It can be seen that there is no clear advantage for small numerosities. Although response times increase from 1 to 4 items, they decrease again for 8 and 32 items. Repeated measures ANOVA in the response times showed that the effect of dot size was significant ($F(1.4, 12.6) = 6.8, P = 0.02$). Trend analysis showed that there was a significant quadratic trend ($F(1, 9) = 87.5, P < 0.001$). This means that the trend in the response times had an inverted U-shape, as can be seen in Fig. 3. There was no significant linear trend. Regression of a linear function to the response times did not yield a significant slope ($P = 0.1, R^2 = 0.5$).

Discussion

Error rates are generally larger than in Experiment 2, indicating that this task was more difficult. This is not surprising given the fact that numerosity information was removed, so there was less information left in the stimuli. However, when numerosity information was absent, subjects were still able to name the different stimuli correctly and there was a significant trend in the response times. This trend was different from the trend that was found in Experiment 2. When numerosity information was removed there was no longer faster or more accurate performance for small numerosities compared to larger numerosities. Consequently, there was no linear trend, showing that there was no increase of the response times from small to large numbers of items. This suggests that black pixels have to be distributed over several disks to enable fast and accurate performance at the first part of the stimulus range. Response times were, however, not constant over the whole range as indicated by the relatively low R^2 value of the linear function. They decrease at both sides of the stimulus range. This was also the case in our haptic study and we have introduced a model to describe this behavior.

Model

It has been shown that response times for judging which of two numbers is larger decreases if the difference between the numbers increases (Moyer and Landauer 1967). This suggests that response times vary with discriminability between numbers. In our paper on haptic numerosity judgement we have introduced a model to describe response times for recognition of a certain stimulus based on discriminability differences between different stimuli (Plaisier et al. 2009). This model describes the pattern of response times only when discriminability follows Fechner's law over the whole range of stimuli. Note that this model describes response times for naming of stimuli that vary in magnitude, not necessarily stimuli differing in numerosity.

However, it is often argued that number representation is similar to magnitude representation. Furthermore, it is possible that numerosity is not accessed directly, but through other co-varying cues like luminance. In our haptic study, the model described the pattern in response times very well when subjects had to label spheres differing in size (i.e. when numerosity information was absent). However, as expected, it could not describe the response times when subjects had to judge varying numbers of spheres in their hand (i.e. when numerosity information was present), indicating that discriminability did not follow Fechner's law over the whole range of numerosities. If indeed similar processes underlie haptic and visual number recognition, then this estimation model should be able to describe the response times from Experiment 3, but not those from Experiment 2 of the present study.

Derivation

Our model assumes that when a presented stimulus has to be recognized and the correct label has to be given, all stimuli in the range are considered weighted according to discriminability between the presented stimulus and each of the other possible stimuli. In accordance with Fechner's law, discriminability is assumed to be proportional to the logarithm of the ratio between the two compared stimuli. The discriminability d between quantities x_1 and x_2 is thus given by:

$$d(x_1, x_2) \propto \left| \log \frac{x_1}{x_2} \right| \quad (3)$$

The total response time is assumed to be inversely proportional to the sum of the discriminabilities. The response time as a function of the presented quantity N can then be described by:

$$T(N) = a + \frac{b}{\sum_{n=i}^j \left| \log \frac{N}{n} \right|} \quad (4)$$

where N is the quantity that is presented, n is an iterator which runs from the smallest quantity in the set (i) to the largest one (j) over all quantities in the set. Free parameters a and b scale the offset and shape of the function. Here, parameter b alone determines the shape of the function, but the average response time over all numerosities in the range (μ) is determined by a combination of a and b :

$$\mu = a + \frac{b \sum_{N=i}^j \frac{1}{\sum_{n=i}^j \left| \log \frac{N}{n} \right|}}{\sum_{n=i}^j 1} \quad (5)$$

Note that this model predicts that response times decrease towards both ends of the stimulus range. For

instance, when the smallest stimulus is presented, there is no smaller one to which it can be compared. Similarly, when the largest stimulus is presented there is no larger stimulus to which it can be compared. Furthermore, if the relative differences between subsequent numerosities are constant, the shape of the function will be symmetrical with the maximum in the middle of the stimulus range. This is illustrated in Fig. 4. In this figure it can also be seen that the predicted response times will depend on the stimulus range that is presented. Because in this model response times are modeled as a function of the presented range it is crucial that data from the whole range are included in the analysis. This was not the case in Experiment 1, where the last stimulus with the largest numerosity was discarded from the analysis because of possible end-effects. The bi-linear model from Experiment 1 does not predict end-effects and to determine the counting slope correctly the last data point should be discarded. The model presented here was fitted to the response times from Experiment 2 and Experiment 3.

Regression analysis

Figure 5a shows the response times for the different numbers of items in Experiment 2. The response times for the different dot sizes from Experiment 3 are shown in Fig. 5b. For both conditions the best fit of the estimation model is represented by the solid line. As can be seen the model cannot describe the data from Experiment 2 ($R^2 = 0.38$) and performs even worse than a linear function. However, it describes the response times from

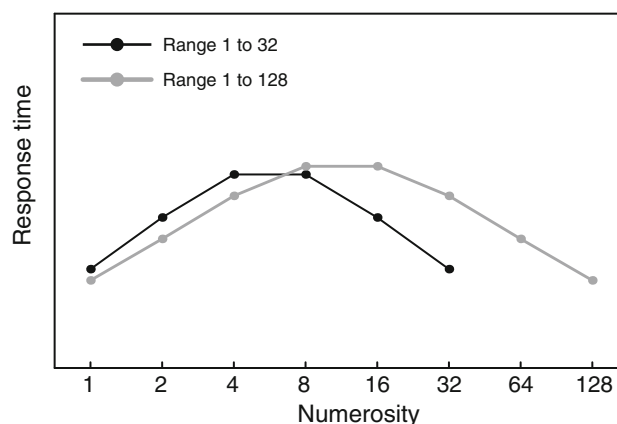


Fig. 4 Predicted pattern in the response times as a function of the number of items. This is a discrete model and only defined at whole numbers. Therefore the predicted response times are indicated by the dots and these were connected for clarity. Response times for a range from 1 to 32 are shown in black, while those for range 1 to 128 are shown in grey. It can be seen that the predicted response times very much depend on the stimulus range. Note that the scaling in the vertical direction is determined by free parameter b . Therefore, the actual response time may be scaled differently comparing both ranges

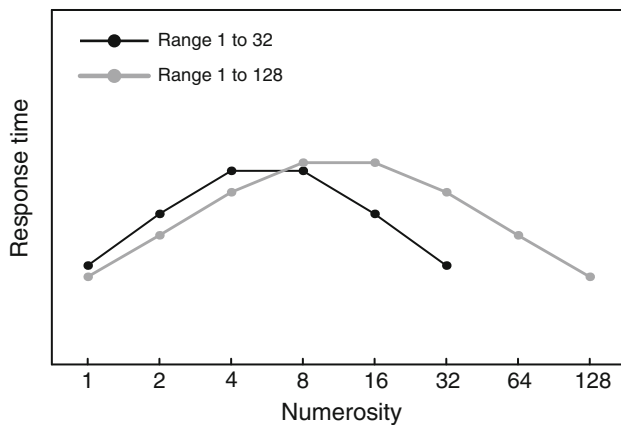


Fig. 5 Response times from Experiment 2a and Experiment 3b with the best fit of the estimation model (*solid line*). The response times from the haptic study are plotted in *grey*. In that case the maximum number was 16. Note the upward shift of the axis

Experiment 3 very well ($R^2 = 0.96$) and much better than a single linear function ($R^2 = 0.5$). The values of the fitting parameters were $b = 2.1$ s and $\mu = 0.7$ s.

Again, the regression analysis was also performed on the data from the single subjects. Averaging the R^2 values from each subject in Experiment 2 yielded $R^2 = 0.009 \pm 0.0009$ (SE). So the model cannot describe the relation between numerosity and response time. This is in agreement with the result from the regression to the response times averaged over subjects. For Experiment 3, this analysis yielded $R^2 = 0.6 \pm 0.09$ (SE), indicating that the model can describe the data in this case. The resulting fitting parameters averaged over subjects were $b = 2.8 \pm 0.3$ s and $\mu = 0.94 \pm 0.04$ s (SE).

Discussion

Our analysis shows that our model describes the response times for Experiment 3, where no numerosity information was present. As expected, it does not describe the data from Experiment 2. In Experiment 2, recognition of small numerosities (<4) was faster than for the larger numerosities. This suggests that discriminability for small numerosities is much larger than for large numerosities even though the relative differences were the same. This is in agreement with what we have reported previously in haptic numerosity judgement. In Fig. 5 the response times from our haptic study are plotted in grey. Note that for clarity the axis for the haptic response times is shifted upwards. It can be seen that the haptic response times correspond relatively well with the response times from the present visual study, although in the haptic case the stimulus range ended at 16 items. In both modalities, faster performance for numerosities from the subitizing range was found than outside this range. In both cases this faster performance

disappeared when stimuli were coded in physical magnitude. This suggests that in both cases response times for the first part of the stimulus range were smaller than for the last part of the stimulus range, but only if numerosity information was present. This indicates that discriminability was better for numerosities from the subitizing range than for larger numerosities. This raises the question whether response times follow a similar pattern as those for magnitude estimation when only numbers larger than the subitizing range are shown.

Experiment 4

In this Experiment we investigated whether discriminability of numbers larger than the subitizing range follows Fechner's law. Therefore, we removed the numerosities in the subitizing regime from the range of numerosities that was used in Experiment 2 and extended the range to larger numerosities. In this experiment we prevented subjects from using other cues like presentation area, density and luminance by using the same method as Izard and Dehaene (2008) recently reported¹ We refer to the area over which the dots were distributed as the 'presentation area' here and not 'occupied area' as Izard and Dehaene did, because this term could be confused with the definition of occupied area as introduced by Allik and Tuulmets (1991). In their definition, 'occupied area' is related to the ratio of empty space to filled space of a display. Filled space is in this model not defined as the sum of the physical area of all dots, but as a region in which these dots lie. This means that the occupied area does not depend on dot size. Therefore, occupied area is related to the spatial distribution of the items in Allik and Tuulmets' occupancy model.

Method

The set-up and task were as described in the "General method" section. Subjects were presented with 8, 16, 32, 64 or 128 dots randomly distributed over the presentation area. They were explicitly told which numbers could be presented. There were three different types of trials. In one third of the trials dot size (0.15° diameter) and presentation area were kept constant (20° diameter). In another third of the trials the presentation area was varied such that dot density was constant for all numerosities (0.15° dot diameter and presentation area ranged from 5.4 to 21.5° diameter). In the last third

¹ This manipulation of the stimuli was not applied in Experiments 1 and 2, to keep the results of these experiments comparable to those of previous studies in which this manipulation was usually not done. Comparison of Experiments 2 and 3 already shows that the pattern in the response times of Experiment 2 are not likely caused by luminance or presentation area estimation only.

of the trials the dot size was varied such that the total luminance was constant for all numerosities (dot diameter varied from 1 to 0.25° and presentation area was 21.7° diameter). All three trial types were interleaved randomly so that only numerosity was a reliant cue in all trials.

Results

Repeated measures ANOVA with numerosity and trial type as factors showed an effect of numerosity ($F(1.3, 12) = 7.8, P = 0.012$) and of trial type ($F(2, 18) = 4.8, P = 0.022$). There was no interaction between both factors ($F(3.2, 29) = 0.98, P = 0.46$) and the quadratic trend was significant ($P = 0.018$). Post-hoc tests (paired t tests with Bonferroni correction) did not show significant differences between the trial types ($P \geq 0.07$). This indicates that there were no significant differences in the shape of the response times for the different trial types. To be certain of this, regression of the estimation model was performed for the three trial types separately. This analysis yielded $b = 5.9$ and $\mu = 1.2$ s for the trials with varying dot sizes, $b = 5.9, \mu = 1.2$ s for the trials with varying presentation area and $b = 6.0, \mu = 1.1$ s for the trials in which presentation area and dot size were constant ($R^2 \geq 0.7$). The lack of significant differences in the shapes of the response times allowed us to collapse the three different trial types. Regression to the data with all trial types collapsed yielded $b = 5.9$ and $\mu = 1.1$ s ($R^2 = 0.8$). Figure 6 shows the response times and error rates averaged over subjects for all numerosities. It can be seen that the response times follow a pattern similar to that found in Experiment 3. The solid line represents regression of the estimation model to the response times averaged over subjects. For comparison, regression of a linear function did, like in Experiment 3, not yield a significant slope ($P = 0.1$) and performed much worse ($R^2 = 0.4$) than our model.

Regression of our model to the single subject response times yielded $R^2 = 0.7 \pm 0.08$ (SE), averaged over all

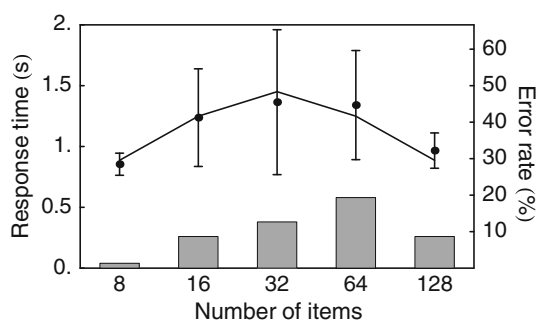


Fig. 6 Response times (dots) and error rates (bars) averaged over subjects from Experiment 4. The solid line represents the best fit of the estimation model to the response times averaged over subjects. Error bars indicate the standard deviation of the single subject means

subjects. The values of the shape parameter and the average response time were $b = 5 \pm 2$ s (SE) and $\mu = 0.9 \pm 0.03$ s (SE), respectively.

Discussion

These results show that our model can indeed describe response times when numerosity information is present when all numerosities are larger than the subitizing range. This indicates that discriminability between subsequent numerosities is constant over this range of numerosities. Note that this conclusion is also supported by the analysis of the three trial types separately and the conclusion does not change depending on whether we collapse the three trial types or not. In Experiment 5 we investigated whether response times for recognition of large numbers are influenced by the presence of numerosities from the subitizing regime in the presented range of numerosities.

Experiment 5

In this experiment we investigated whether numerosities from the subitizing regime are taken into consideration during the estimation of larger numerosities. If they are, then adding them to the numerosity range should yield the inverted U-shaped pattern from Experiment 4, but now symmetrical around 8 and 16 (the middle of the range). However, if they are not taken into consideration, then the pattern in the response times should be the same as found in Experiment 4. In this last case we can conclude that small numbers are not taken into consideration or discarded very fast when a large number is presented.

Method

Subjects were presented with 1, 2, 4, 8, 16, 32, 64 or 128 dots randomly distributed over the presentation area. Again subjects were explicitly told which numbers could be presented. Luminance and dot density cues were removed as described in the Method section of Experiment 4. In the trials where dot density was constant for all numerosities, the presentation area now ranged from 1.9 to 21.5° diameter and in the constant luminance trials the dot size ranged from 2.8 to 0.25° diameter.

Results

Repeated measures ANOVA with numerosity and trial type as factors showed an effect of numerosity ($F(1.9, 17) = 23.4, P < 0.0001$), but not of trial type ($F(2, 18) = 2.6, P = 0.099$). Therefore, the data from the three different types of trials were collapsed. Response times and

error rates averaged over subjects are shown in Fig. 7. It can be seen that from numerosity 8 and larger the response times follow a similar pattern as found in Experiment 4. The estimation model was fitted to the response times averaged over subjects for different numerosity intervals. The interval over which the quality of the fit is best, indicates the range of numerosities that is included in the estimation process. As was shown earlier, the shape of the model depends on the range of numerosities (Fig. 4). There were six intervals ranging from 1 to 128, 2 to 128 and so on to the interval from 32 to 128. The R^2 values that were found were 0.3, 0.5, 0.8, 0.9, 0.7 and 0.2, respectively. The optimum in the quality of the fit was thus found over the interval from 8 to 128, i.e. all numerosities well outside the subitizing regime. Regression of the model over this interval is represented with the solid black line in Fig. 7. The value of the shape parameter and the average response times were found to be $b = 3.5$ s and $\mu = 1.2$ s, respectively. For comparison, regression of a linear function was performed for the whole range of stimuli and over the interval from 8 to 128 separately. Over the whole range the resulting R^2 value was 0.15 and for the interval from 8 to 128, R^2 was 0.12. This shows that our model describes the data much better than a linear function.

Also, regression to the single subjects' data was performed. This yielded on average $R^2 = 0.6 \pm 0.1$ (SE), so the model fitted the data well. The shape parameter and average response time were found to be $b = 4 \pm 1$ s (SE) and $\mu = 1.2 \pm 0.8$ s (SE), respectively.

As the same subjects participated in both Experiments 4 and 5 and they performed the experiment in counterbalanced order, the fitting parameters were compared between the experiments. Paired-samples t tests yielded no significant differences ($P \geq 0.07$) between the experiments for both parameters.

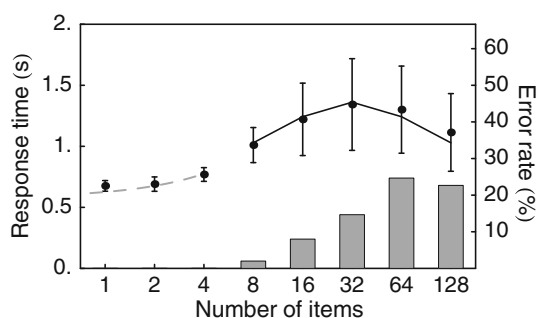


Fig. 7 Response times (dots) and error rates (bars) averaged over subjects from Experiment 5. The solid black line represents the best fit of the estimation model to the response times. The grey dashed line is the first linear part from the fit of the bilinear function to the data from Experiment 1 plotted on an logarithmic scale. Error bars indicate the standard deviation of the single subject means

The dashed grey line in Fig. 7 is the result from the fit for numerosities in the subitizing regime from experiment 1, re-plotted on a logarithmic scale. Because of the logarithmic scaling, the linear function is now curved. It can be seen that the line fits also the response times from this experiment, even though different subjects participated in both experiments. This shows that the response times for numerosities in the subitizing regime were not affected by the difference in the presented numerosities between this experiment and Experiment 1.

Discussion

The results show that adding numerosities from the subitizing regime did not significantly change the response times for numerosities outside the subitizing range. The pattern in the response times was symmetrical around 32 items, which was the middle numerosity between 8 and 128 (i.e. the numerosities outside the subitizing regime). This indicates that numerosities from the subitizing regime were not taken into consideration when numerosities outside the subitizing range were presented. Furthermore, the response times in the subitizing range were comparable to those found in Experiment 1. This indicates that the subitizing process was relatively unaffected by the differences between the numerosity ranges used in Experiment 1 and Experiment 5. These results show that numbers from the subitizing range are not taken into consideration or were discarded very fast when a numerosity outside the subitizing range was shown and vice versa.

General discussion

The results from Experiments 1, 2 and 3 are in agreement with the results from our haptic study (Plaisier et al. 2009). Note that the stimuli differ in many ways between the haptic study and the present visual study. In the haptic case, spheres were grasped and could be actively rearranged in the hand. In vision there is no such active control over the positions of the dots. In the case of vision, on the other hand, pattern recognition may play a role. Pattern recognition has been suggested as an explanation for subitizing (Mandler and Shebo 1982). Pattern recognition does not seem applicable to the haptic case as the positions of the spheres were not fixed. Moreover, pattern recognition is not likely to have played a role in the study on tactile subitizing where varying numbers of fingers were stimulated (Riggs et al. 2006). The fact that despite these differences, numbers up to three or four are recognized faster and more accurately than larger numbers in vision as well as haptics suggests that the underlying reason may be the same in both modalities. This has interesting implications for the

possible processes underlying numerosity judgement, as these should be processes that extend across both modalities. Consequently, pattern recognition is not a very likely explanation. The idea that number representation is modality independent is not unlikely. Using brain imaging (fMRI) it has been shown that there is cross-notational (arabic numerals and dot patterns) adaption to number (Piazza et al. 2007). Furthermore, it has been suggested that representation of number and physical magnitude is shared (Walsh 2003). This suggests that number is encoded in an abstract fashion and representation independent. This representation might very well be modality independent.

From Experiments 4 and 5 it is clear that numerosities from the subitizing range are not taken into consideration when numerosities larger than the subitizing range are shown. This is in line with the idea that subitizing means that subjects almost instantaneously know which numerosity is presented. This does not only mean that subjects perform practically error-free in the subitizing regime, they also know very quickly whether or not the presented numerosity can be subitized. The results from Experiments 2 and 5 both show that even if the relative spacing between subsequent numerosities is large over the whole numerosity range, there is an advantage for judgement of small numerosities. So constant relative magnitude differences between the numerosities do not enable subitizing for larger numerosities. It was mentioned before that pattern recognition is also not a likely explanation. Still, it seems that numerosities from the subitizing regime are recognized as ‘subitizable’ very efficiently. It has been shown that numerosities from the subitizing range are rated as more dissimilar than numerosities from outside that range (Logan and Zbrodoff 2003). This would explain why adding numerosities from the subitizing regime did not affect the response times for recognition of larger numerosities (Experiment 5) much. Now the question arises of what enables this fast recognition of small numerosities?

An explanation for the subitizing mechanism that does not involve discriminability or pattern recognition is based on visual indexing theory (see Pylyshyn (2001) for a review). According to this theory humans can refer to an item without linking it to a specific feature like position. Such an indexing system can be used for directing attention to certain objects or for motor actions like eye or grasping movements towards objects. From visual tracking studies, it was found that subjects can track up to 5 items simultaneously and it is hypothesized that the number of items that can be referred to simultaneously in this way is limited to 5 (Pylyshyn and Storm 1988). This idea can also be used to explain why numerosities smaller than 5 can be judged faster and more accurately than larger numerosities (Trick and Pylyshyn 1994). The idea that indexing is used for directing attention could be easily extended to the haptic

modality. Although there is no prior evidence that a process like haptic indexing exists, it is not unlikely that indexing also occurs in the haptic modality.

In conclusion, we have shown that there is an advantage for judging of small numerosities (<4) over large numerosities even if the relative differences between subsequent stimuli is a factor of 2 over the stimulus range. This advantage was not mediated by recognition of the numerosities through judgement of density, presentation area or luminance. Furthermore, the faster performance for the smallest stimuli in the range disappeared when numerosity information was removed. This supports the idea that subitizing does not reflect very accurate estimation mediated through large differences between subsequent numerosities. Furthermore, we would like to propose that similar processes underly haptic and visual numerosity judgement.

Acknowledgments This research was supported by the Netherlands Organisation for Scientific Research (NWO).

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Agrillo C, Dadda M, Bisazza A (2007) Quantity discrimination in female mosquitofish. *Anim Cogn* 10:63–70
- Akin O, Chase W (1978) Quantification of three-dimensional structures. *J Exp Psychol Hum Percept Perform* 4:397–410
- Allik J, Tuulmets T (1991) Occupancy model of perceived numerosity. *Percept Psychophys* 49:303–314
- Atkinson J, Campbell FW, Francis MR (1976) The magic number 4 ± 0 : a new look at visual numerosity judgments. *Perception* 5:327–334
- Balakrishnan JD, Ashby FG (1992) Subitizing: magical numbers or mere superstition? *Psychol Res* 54:80–90
- Beran MJ, Tagliabue LA, Flemming TM, James FM, Washburn DA (2006) Nonverbal estimation during numerosity judgements by adult humans. *Q J Exp Psychol A* 59:2065–2082
- Boysen ST (1997) Representation of quantities by apes. *Adv Study Behav* 26:435–462
- Camos V, Tillmann B (2008) Discontinuity in the enumeration of sequentially presented auditory and visual stimuli. *Cogn Tech Work* 107:1135–1143
- Cordes S, Gelman R, Gallistel CR, Whalen J (2001) Variability signatures distinguish verbal from nonverbal counting for both large and small numbers. *Psychon Bull Rev* 8:698–707
- Dehaene S, Dehaene-Lambertz G, Cohen L (1998) Abstract representations of numbers in the animal and human brain. *Trends Neurosci* 21:355–361
- Gallistel CR, Gelman R (1992) Preverbal and verbal counting and computation. *Cogn Tech Work* 44:43–74
- Gallistel CR, Gelman R (2000) Non-verbal numerical cognition: from reals to integers. *Trends in Cogn Sci* 4:59–65
- Izard V, Dehaene S (2008) Calibrating the mental number line. *Cogn Tech Work* 106:1221–1247
- Kaufman E, Lord M, Reese T, Volkman J (1949) The discrimination of visual number. *Am J Psychol* 62:498–525

- Logan GD, Zbrodoff NJ (2003) Subitizing and similarity: toward a pattern-matching theory of enumeration. *Psychon Bull Rev* 10:676–682
- Logie RH, Zucco GM, Baddeley AD (1987) Interference with visual short-term-memory. *Bull Br Psychol Soc* 40:A107
- Mandler G, Shebo BJ (1982) Subitizing: an analysis of its component processes. *J Exp Psychol Gen* 111:1–22
- Moyer RS, Landauer TK (1967) Time required for judgements of numerical inequality [47]. *Nat Biotechnol* 215:1519–1520
- Nieder A, Miller EK (2003) Coding of cognitive magnitude: compressed scaling of numerical information in the primate prefrontal cortex. *Neuron* 37(1):149–157
- Nieder A, Freedman DJ, Miller EK (2002) Representation of the quantity of visual items in the primate prefrontal cortex. *Sci Agric* 297:1708–1711
- Oyama T, Kikuchi T, Ichihara S (1981) Span of attention, backward masking, and reaction time. *Percept Psychophys* 29:106–112
- Pepperberg IM (2006) Grey parrot numerical competence: a review. *Anim Cogn* 9(4):377–391
- Piazza M, Pinel P, Le Bihan D, Dehaene S (2007) A magnitude code common to numerosities and number symbols in human intraparietal cortex. *Neuron* 53(2):293–305
- Plaisier MA, Bergmann Tiest WM, Kappers AML (2009) One, two, three, many—subitizing in active touch. *Acta Psychol* 131:163–170
- Pylyshyn ZW (2001) Visual indexes, preconceptual objects, and situated vision. *Cogn Tech Work* 80:127–158
- Pylyshyn ZW, Storm RW (1988) Tracking multiple independent targets: evidence for a parallel tracking mechanism. *Spat Vis* 3:179–197
- Revkin SK, Piazza M, Izard V, Cohen L, Dehaene S (2008) Does subitizing reflect numerical estimation? *Psychol Sci* 19:607–614
- Riggs KJ, Ferrand L, Lancelin D, Fryziel L, Dumur G, Simpson A (2006) Subitizing in tactile perception. *Psychol Sci* 17:271–272
- Roberts WA, Roberts S, Kit KA (2002) Pigeons presented with sequences of light flashes use behavior to count but not to time. *J Exp Psychol Anim Behav Process* 28(2):137–150
- Ross J (2003) Visual discrimination of number without counting. *Perception* 32:867–870
- Ten Hoopen G, Vos J (1979) Effect of numerosity judgement of grouping of tones by auditory channels. *Percept Psychophys* 26:374–380
- Trick LM (2005) The role of working memory in spatial enumeration: patterns of selective interference in subitizing and counting. *Psychon Bull Rev* 12(4):675–681
- Trick LM (2008) More than superstition: differential effects of featural heterogeneity and change on subitizing and counting. *Percept Psychophys* 70:743–760
- Trick LM, Pylyshyn ZW (1993) What enumeration studies can show us about spatial attention: evidence for limited capacity preattentive processing. *J Exp Psychol Hum Percept Perform* 19:331–351
- Trick LM, Pylyshyn ZW (1994) Why are small and large numbers enumerated differently? A limited-capacity preattentive stage in vision. *Psychol Rev* 101:80–102
- Van Oeffelen MP, Vos PG (1982) A probabilistic model for the discrimination of visual number. *Percept Psychophys* 32:163–170
- Walsh V (2003) A theory of magnitude: common cortical metrics of time, space and quantity. *Trends Cogn Sci* 7:483–488
- Watson DG, Maylor EA, Bruce LAM (2007) The role of eye movements in subitizing and counting. *J Exp Psychol Hum Percept Perform* 33:1389–1399
- West RE, Young RJ (2002) Do domestic dogs show any evidence of being able to count? *Anim Cogn* 5(3):183–186
- Whalen J, Gallistel CR, Gelman R (1999) Nonverbal counting in humans: the psychophysics of number representation. *Psychol Sci* 10:130–137