A New Approach to the Study of Subitizing as Distinct Enumeration Processing

Tony Simon
School of Psychology
Georgia Institute of Technology
Atlanta, GA 30332-0170
(404) 894-2681
tonys@zunow.gatech.edu

Angel Cabrera
School of Psychology
Georgia Institute of Technology
Atlanta, GA 30332-0170
angel@psy.gatech.edu

Reinhold Kliegl
Max Planck Institute for Human Development & Education
Lentzeallee 94, D-1000
Berlin 33 Germany
kliegl@vax1.mpib-berlin.mpg.dbp.de

Abstract*

This paper presents a new methodology for examining the phenomenon of subitizing. Subjects were presented with a standard numerosity-detection task but for a range of presentation times to allow Task-Accuracy Functions to be computed for individual subjects. The data appear to show a continuous change in processing for numerosities from 2 to 5 when the data are aggregated across subjects. At the level of individual subjects, there appear to be qualitative shifts in enumeration processing after 3 or 4 objects. The approach used in this experiment may be used to test the claim that subitizing is a distinct enumeration process that can be used for small numbers of objects.

Introduction

The phenomenon now known as *subitizing* has been reported since the beginning of experimental psychology (Jevons, 1871; Wundt, 1896). Subitizing refers to the ability of subjects to identify the numerosity of small collections of discrete objects very rapidly and without error. It has been suggested that this performance pattern reveals basic characteristics of the perceptual/representational system (Pylyshyn, 1989) and has also been implicated as the foundation for children's conceptual development in the field of number (Simon, Klahr & Newell, 1992). Yet, there is some debate about how many objects can be processed by subitizing. The earliest studies claimed 6 objects could be enumerated but more commonly a limit of 3 to 4 objects is proposed (e.g. Atkinson, Campbell & Francis, 1976).

Subitizing experiments are usually one of two types. In one, subjects are presented with visual stimuli for a single, brief exposure and asked to report the number of objects seen. In the other, presentation is terminated when the subject makes a response. The different designs produce similar results (see Mandler & Shebo, 1982 for a review). Subjects exhibit subitizing for small numerosities followed by other, more time-consuming and/or errorful processes for larger ones. These include estimation, counting and perceptual grouping.

The most typical characterization of the phenomenon is the presentation of a single reaction-time curve split into two regions. For adults, the curve for subitizing up to 3 or 4 objects is shallow and rises at about 50 milliseconds per item. Almost no errors are made in this range. The segment for more than 4 or 5 objects tends to rise steeply at about 300 milliseconds per item with errors starting at around 20% and increasing. Determining where to split the curve is done by looking for the appearance of a quadratic trend in the reaction time data. Where the break occurs is taken to indicate the capacity limit for subitizing.

Despite the fact that most studies on rapid enumeration report data consistent with the pattern described above, there is still considerable controversy over whether subitizing really exists as a distinct phenomenon and, if it does, whether it is a serial or parallel process.

Information processing accounts have been constructed in an attempt to specify the details of the subitizing phenomenon. Klahr & Wallace's (1976) production-system models described subitizing in template-matching terms. The subitizing slope is a function of serially matching templates from zero to N items onto the input at a constant rate. Subitizing limit is a function of the processor rate and the decay of the visual store that holds the to-be-matched items.

* Angel Cabrera is supported by a Fulbright/MEC scholarship from the Institute of International Education & the Spanish Ministry of Education and Science.

Mandler & Shebo (1982) claim three processes are responsible for the typical data pattern. The first is parallel pattern-recognition of the canonical patterns that arrays in the 1 to 3 range exhibit. (One is always a singleton, 2 always a line and 3 either a line or a triangle). This predicts that the oft-seen subitizing slope should, in fact, be flat. Thus they dispute the existence of subitizing as it is usually characterized. Next is "a response to arrays of 4 to 6 or 7 that is based on mental counting" and this is proposed to account for the steeper "post-subitizing" slope. Finally, "an estimating response for arrays larger than 6 (1982, p.1)" is presented as the reason many studies show a leveling-off of slopes for larger numbers. Thus, there is still considerable debate over both the existence and nature of subitizing.

Other investigations have focused on statistical models fit to the reaction time data. Balakrishnan & Ashby (1991) contrast a bilinear model (which suggests the existence of a distinct subitizing process e.g. Chi & Klahr, 1975) with a log-linear model (which suggests a continuous exponential function of reaction time with numerosity e.g. Kaufman et al, 1949). They conclude that neither model adequately explains the relationship between reaction time and stimulus numerosity.

We contend that the lack of theoretical consensus for this apparently robust phenomenon is largely due to the methodological and analytical conventions of the paradigm. By aggregating reaction time data from individual subjects, distinctions between subitizing and other enumeration strategies will be blurred (Siegler, 1987). This is especially true if, for a given numerosity, some subjects are contributing a lot of variance having switched to slow, errorful processing while others are contributing little variance by continuing to use the fast, accurate subitizing process. The difficulty in making a clear distinction between enumeration processes in a subject population serves to fuel the debate over the existence and nature of subitizing.

In this paper we present an alternative methodology which, in the long run, can provide a strict test of the claim that subitizing is a distinct mode of enumeration processing. It also indicates the maximum number of objects that each individual studied can enumerate in this way. Our approach is to determine Time-Accuracy Functions (TAFs) for individual subjects (Kliegl, Mayr & Krampe, 1993) on an otherwise standard subitizing task. Here, presentation time is varied and response curves for a given problem are plotted against the whole range of times. This yields a TAF for each numerosity. If subjects produce a similar function over several numerosities then we take that to indicate use of the same processing. Thus a switch in process is predicted to produce a change in the nature of the function.

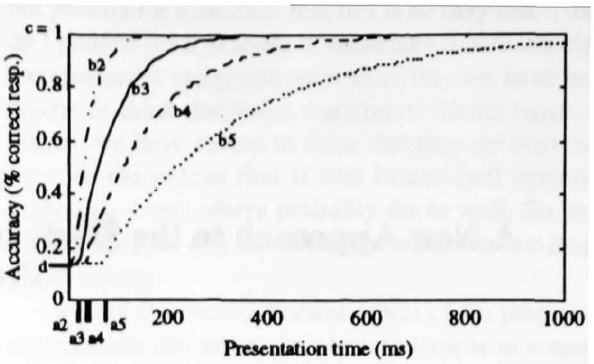

Figure 1. Example functions for the exponential model

An example set of functions for numerosities 2 to 5 is presented in Figure 1. We modeled the relation between time and accuracy with reference to the negatively accelerated exponential function given in Equation 1

$$p = d + (c - d) * (1 - e^{-\frac{t-a}{b}}) \qquad (1)$$

where p indicates the probability of a correct response, t the available presentation time, a the intercept of the function on the time-axis, b the steepness of the curve, c the asymptotic maximum, and d the chance level of performance. Parameter a indicates the amount of time necessary to initialize relevant cognitive processes; parameter b specifies the constant proportional reduction of error probability for increases in presentation time; parameter c was fixed at 1.0 because the enumeration task is simple enough for everybody to yield perfect performance with sufficient time; parameter d was fixed at 1/7 because participants had to respond with one of seven response keys (2 to 8).

This function has been used widely in psychological research, for example in the identification of test tones as a function of the silent interval between test and masking tone (Massaro, 1970), for the accumulation of information in iconic memory (Loftus, Duncan, & Gehrig, 1992), for describing speed-accuracy tradeoffs (Wickelgren, 1977), and for visual search tasks, episodic memory, and complex reasoning tasks (Kliegl, Mayr, & Krampe, 1993). All approaches share the conceptual interpretation that it becomes increasingly difficult to improve performance relative to the level of accuracy reached.

The current experiment tested the identification of specific numerosities from 2 to 8 objects. The predictions are that, as shown in Figure 1, for higher numerosities the function will shift to the right on the x-axis (i.e. parameter a will increase) and the slope of the function will decrease (i.e. parameter b will increase). If parameter b changes by a constant factor across all numerosities, this would constitute evidence for a continuous process of enumeration; if there is a qualitative shift somewhere, this should be visible by a jump in the ratio of the neighboring values of the b parameter. Importantly, the approach presented here allows us to look for this shift at the level of group as well as individual functions.

Method

Subjects

Twenty undergraduate students at Georgia Institute of Technology participated in the experiment for extra credit. There were 10 males and 10 females.

Stimuli

Rows of 2 to 8 lower case Helvetica 9 point typeface letter O's without spaces between them were constructed for presentation. Pilot testing had indicated that stimuli this small were required to induce some errorful performance for very small numbers, even at the shortest presentation times.

Each row was presented in the center of a computer screen. A fixation arrow pointed to the middle of where the central letter O (for odd numbers) was to be displayed, or in between where the two most central O's would be (for even numbers). The longest row (N=8) was 5/8" in length. A mask of 16 identical letter O's centered at the fixation point was also used. The mask was 1 1/8" in length. Stimuli were presented approximately at eye level 30" from the subject so that visual angle for stimuli was less than 2 degrees. All stimuli were presented on a high-resolution 13" Apple Macintosh™ RGB monitor switched to black and white mode.

Procedure

The experiment was run on an Apple Macintosh™ IIfx using Cedrus Superlab™ 1.5 software. There were 12 stimulus presentation times. These were 34, 51, 67, 84, 100, 150, 200, 250, 300, 500, 750 & 1000 milliseconds. Seventy-two stimulus-duration pairs were created. Stimuli of N= 2 and 3 were presented for durations from 34 to 300 milliseconds. Beyond that time pilot subjects had shown perfect performance. Stimuli of N= 7 and 8 were presented for durations from 84 to 1000 milliseconds. Pilot data showed numerosity detection to be extremely difficult at 84 milliseconds. Stimuli of N= 4, 5 and 6 were shown for all twelve presentation times. There were 24 blocks of randomly ordered trials providing a total of 1728 responses from each subject. Treatments were randomly dispersed among all blocks and treatment blocks were randomly re-ordered for each subject.

For each trial the subject saw a fixation arrow presented on the screen for 750 milliseconds and was told that the row of objects would appear centered at the tip of the arrow. Then the stimulus was presented for its preselected time, followed by the mask. At this point the subject could make his or her response, which was to type the number of objects that had appeared before the mask. At the end of each block of 72 trials the program halted and allowed the subject to restart whenever he or she was ready. At the end of every 8 blocks (576 trials) the program requested that the subject take a two-minute break before continuing.

Subjects were brought into the lab individually, the experiment was explained to them and then they were given 72 practice trials. Here a digit was presented in the center of the screen and the subject had to type the same number on the keyboard to remove it. This was done to give the subject practice in locating the keys for the numbers 2 through 8 and reduce search during the experiment. Finally, the subject completed 5 demonstration trials to ensure they understood the procedure.

Results

For every subject, 24 responses were recorded for each of the 72 different stimulus-duration pairings. As described in the example function, probability correct scores were calculated to allow examination of the effect of duration on each subject's ability to reach correct detection for set sizes from N= 2 to 8. These values were then used to fit the negatively accelerated exponential function described above using iterative non-linear regression. Functions could only be fit accurately for numerosities 2 through 5, therefore only these will be reported. This is because a level of approximately 40% correct responses is required to compute the function and such a level of performance was rarely achieved for numerosities 6 though 8.

A minimum chi-square statistic was used as an estimator to take into account the change in variance across accuracies. For each cell of the numerosity (2 to 5) by presentation time by type of response (correct/incorrect) contingency table, we computed $\frac{(no-ne)^2}{ne}$ where no is number of observed and ne the number of expected responses given the exponential equation (Eq. 1). Parameters were sought that minimized the sum of these terms using the CNLR module of SPSS-X. Parameters a and b were constrained to be equal or increase with numerosity;

Table 1. Values of parameters a and b for each subject for numerosities N=2-5.

	parameter a				parameter b					parameter a				parameter b			
S#	2	3	4	5	2	3	4	5	S#	2	3	4	5	2	3	4	5
1	13	51	56	132	23	42	173	173	11	0	0	75	75	31	71	71	227
2	0	0	47	47	36	63	107	291	12	0	15	15	16	57	57	57	257
3	8	8	8	8	58	72	119	407	13	11	20	20	20	62	120	268	513
4	30	30	30	30	24	56	57	201	14	26	28	34	39	48	54	209	425
5	15	19	59	59	27	85	140	423	15	0	28	38	39	0	58	95	158
6	17	18	18	100	26	73	92	314	16	19	29	64	64	23	34	118	169
7	23	40	40	119	16	57	85	223	17	0	0	35	83	0	53	71	92
8	0	0	0	0	75	103	217	387	18	16	27	27	27	32	44	106	192
9	25	25	25	25	32	125	193	512	19	0	0	0	145	40	75	202	231
10	0	16	22	142	0	45	201	309	20	0	18	18	116	51	53	150	174

that is we assumed that an increase in the number of objects to be enumerated requires at least the same amount of cognitive processing as the lower number of objects.

The overall goodness of fit across 20 subjects was significant but still acceptable given the large number of observations (minimum chi-square = 1127 for 680 degrees of freedom, $p<.01$); note the ratio of chi-square and degrees of freedom is less than two. Table 1 presents the values of parameters a and b for each subject for numerosities N= 2-5 computed from the non-linear regression. The values were multiplied by 100 and rounded to a whole number.

It can be clearly seen from the table that both parameters a and b increased with numerosity. This is also demonstrated in the individual time-accuracy functions shown in Figure 2. An important result is that the b parameter increased by a constant factor of two across numerosity. If we aggregate across subjects the values are 30, 70, 140 & 280 for numerosities N= 2-5. Thus, there is no evidence for a discontinuity in processing between two and five objects. This suggests that, if we assume that subjects are subitizing 2 objects, then they appear to be processing 5 objects in the same way. The functions displayed in Figure 1 were actually based on the group averages of estimates for parameters a and b. However, examination of individual data suggests that processing has changed in most subjects by this point. This divergence may point to the very weakness of aggregating data that inspired the current study.

We can illustrate this by examining the data in more detail. Table 1 shows values of the a and b parameters for each numerosity across subjects. We assume that the point at which a subject switches processing is indicated by a shift in the ratio between neighboring numerosities. On the basis of previous research, the question appears to be whether the shift occurs after 3 or 4 objects. We can examine this issue at the individual level by computing the ratios between b(N=4)/b(N=3) and b(N=5)/b(N=4). For 10 subjects the second ratio was larger than the first, indicating a shift after 4 objects. For 9 subjects the first ratio was larger than the second, indicating a shift after 3 objects, while for 1 subject the ratios were equal.

Figure 2 presents the TAFs of four subjects (#s 1, 6, 7 & 16). Each plot depicts the exponential functions of accuracy (% correct) for numerosities 2 through 5 for presentation times from 34 to 1000 ms. Values of each subject's parameter b provide much information about his or her performance. They can be thought of as an indication of how quickly the subject approaches perfect detection of a given numerosity. Low values of b, and their associated steep curves, indicate that small increments to initial presentation time enable the subject to accurately detect the numerosity being displayed.

The plots show that the time needed to reach asymptote increases systematically as numerosity increases. This is directly related to the rise in the value of the b parameter (see Table 1). It can be observed that subjects 1 & 16 reach asymptote very quickly for N= 2 & 3 but require very large increments of time before correctly identifying numerosities of N= 4 or 5. Subjects 6 & 7 show a similar pattern for N= 2 to 4 but require large time increments to approach correct detection of N= 5.

The pattern of error-free detection within short presentation times indicates the use of a fast, accurate enumeration process, i.e. subitizing. Its use is only apparent for small numbers. For larger numbers, errors are made at all durations. These errors decrease with longer presentation times suggesting they would eventually disappear. Such a pattern indicates more time-consuming and errorful strategies such as estimation or counting. Thus subjects 1 & 16 appear to subitize to N= 3 while subjects 6 & 7 appear to subitize to N= 4.

This novel methodology suggests that, on rapid enumeration tasks, subjects may exhibit a distinct

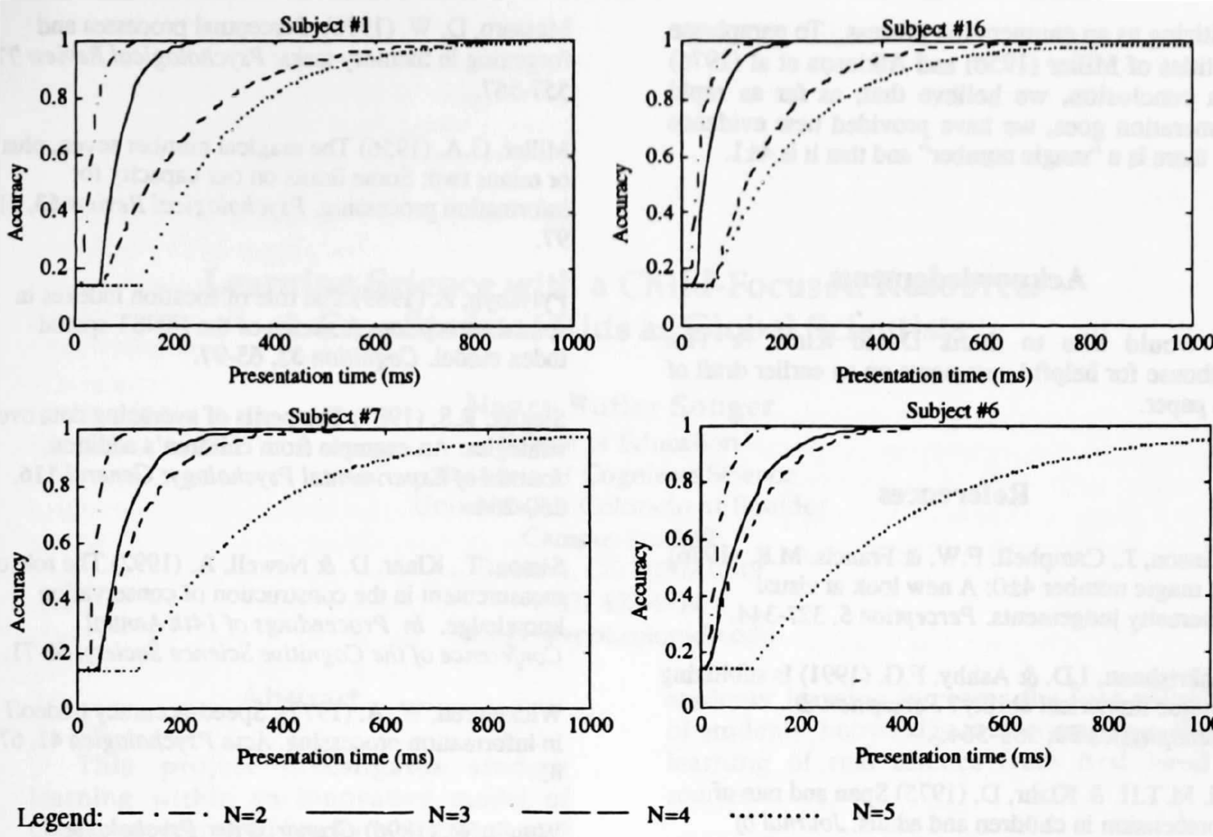

Figure 2. Plots of exponential functions for subjects 1, 6, 7 & 16 for numerosities N= 2 to 5.

processing pattern, apparently with a capacity of N= 3 or 4. The ratios (i.e. the shifts noticeable in Table 1) are quite suggestive of qualitative shifts. Unfortunately, for the present set of data we can not determine the reliabilities of these individual shifts. We expect this to be possible with data collected from an extended testing schedule.

Conclusions

The evidence presented here is consistent with the view that subjects employ a single fast, accurate process when rapidly enumerating up to four or five objects. For larger collections they apparently use a range of slower, less accurate processes. The novel methodology we have used enables detailed examination of the subitizing phenomenon at the individual subject level. The data it has produced appears to be highly consistent with other characterizations of the subitizing phenomenon, even though the analyses are novel and are undergoing refinement. There is some indication that enumeration performance shows a discontinuity in processing at around 3 to 4 objects. This supports the view that subitizing does exist as a distinct process for rapid enumeration and that it is limited to small set sizes. For larger set sizes it seems clear that subjects must rely on a range of slower and less accurate strategies such as counting, estimation and guessing.

Our conclusions share some features of both models evaluated by Balakrishnan & Ashby (1991). Our data are in line with the conclusions of Klahr and his colleagues that a distinct subitizing process exists, but it is still necessary to provide statistical tests. We have also shown that an exponential function serves as a good model for rapid enumeration processing. It is similar to Kaufman et al.'s (1949; see also Balakrishnan & Ashby, 1991) log-linear model but applied to presentation times required for different levels of accuracy rather than response latencies. Its fit to smaller rather than larger numerosities for the range of presentation times examined can be taken as preliminary evidence that a single process is used by all subjects in response to small numbers whereas each subject may use a range of strategies as the set size increases. In future research this issue needs to be examined with the use of longer presentation times for larger numerosities.

The real strength of our findings, although preliminary, is that they provide converging evidence from a novel methodology for the existence of the subitizing phenomenon. Much still remains to be done to refine the technique and to carry out further studies that reveal the precise processing details of subitizing as an enumeration process. To paraphrase the titles of Miller (1956) and Atkinson et al (1976) as a conclusion, we believe that, as far as rapid enumeration goes, we have provided new evidence that there is a "magic number" and that it is 4±1.

Acknowledgments

We would like to thank David Klahr & Tim Salthouse for helpful comments on an earlier draft of this paper.

References

Atkinson, J., Campbell, F.W. & Francis, M.R. (1976) The magic number 4±0: A new look at visual numerosity judgements. *Perception* **5**, 327-344.

Balakrishnan, J.D. & Ashby, F.G. (1991) Is subitizing a unique numerical ability? *Perception & Psychophysics* **50**, 555-564.

Chi, M.T.H. & Klahr, D. (1975) Span and rate of apprehension in children and adults. *Journal of Experimental Child Psychology* **19**, 434-439.

Jevons, W.S. (1871) The power of numerical discrimination. *Nature* **3**, 281-282.

Kaufman, E.L., Lord, M.W., Reese, T.W. & Volkmann, J. (1949) The discrimination of visual number. *American Journal of Psychology* **30**, 498-525.

Klahr, D. & Wallace, J.G. (1976) *Cognitive Development: An Information Processing View.* Hillsdale, NJ: Erlbaum.

Kliegl, R., Mayr, U. & Krampe, R.T. (1993) Time-accuracy functions for determining process and person differences. An application to cognitive aging. Manuscript submitted for publication.

Loftus, G.R., Duncan, J., & Gehrig, P. (1992). On the time course of perceptual information that results from a brief visual presentation. *Journal of Experimental Psychology: Human Perception and Performance* **18**, 530-549.

Mandler, G. & Shebo, B.J. (1982) Subitizing: An analysis of its component processes. *Journal of Experimental Psychology: General* **111**, 1-22.

Massaro, D. W. (1970). Perceptual processes and forgetting in memory tasks. *Psychological Review* **77**, 557-567.

Miller, G.A. (1956) The magical number seven, plus or minus two: Some limits on our capacity for information processing. *Psychological Review* **63**, 81-97.

Pylyshyn, Z. (1989) The role of location indexes in spatial perception: A sketch of the FINST spatial-index model. *Cognition* **32**, 65-97.

Siegler, R.S. (1987) The perils of averaging data over strategies: An example from children's addition. *Journal of Experimental Psychology: General* **116**, 250-264.

Simon, T., Klahr, D. & Newell, A. (1992) The role of measurement in the construction of conservation knowledge. *In Proceedings of 14th Annual Conference of the Cognitive Science Society*, 66-71.

Wickelgren, W. A. (1977). Speed-accuracy tradeoff in information processing. *Acta Psychologica* **41**, 67-85.

Wundt, W. (1896) *Grundriss der Psychologie.* Leipzig: Engelmann.