Evidence for Subitizing as a Stimulus-Limited Processing Phenomenon

Tony J. Simon
School of Psychology
Georgia Institute of Technology
Atlanta GA 30332-0170
tony.simon@psych.gatech.edu

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

Abstract

We present an experiment where subject's subitizing performance for linear dot arrays was analyzed using Differential Time Accuracy Functions. This technique uses accuracy and reaction time data to decompose overall response latency into stimulus-limited and post-stimulus processing. Our results show that subitizing is a phenomenon produced by the effects of increased numerosity on stimulus-limited processes alone. They also suggest that the familiar guessing strategy for the largest arrays in reaction time measures of subitizing results from a reduction in post-stimulus processing. Subjects appear to extract the perceptual characteristics of all arrays but presumably fail for the largest and therefore default to guessing. Existing theories of subitizing are evaluated in light of these results.

Introduction

The phenomenon known as subitizing has been implicated as a foundational capability in the development of numerical and mathematical competence (Geary, 1995; Simon & Klahr, 1995). Yet, despite its investigation by psychologists for more than 100 years (e.g. Jevons, 1871) a definitive information processing account remains to be developed. Subitizing describes the ability of adults and children to very rapidly and accurately enumerate a small number of discrete entities. The number of items is typically in the range of 3-5. Beyond that number, enumeration proceeds at a much slower rate and with considerably more errors (Atkinson, Campbell & Francis, 1976; Chi & Klahr, 1975; Svenson & Sjoberg, 1983). Thus, a satisfactory account of subitizing must adequately explain 2 aspects of the phenomenon. One is the reason for the very rapid rate of processing for small numerosities, which is evident in a slope of around 40 to 100 milliseconds per item. The other is the limit or span of subitizing, which is defined by a sharp discontinuity in performance that occurs when subjects reach the maximum number of items they can enumerate in this way.

Despite the lack of consensus on a detailed explanation of subitizing, there is considerable agreement on a higher level of analysis. Subitizing is assumed, if only implicitly, to comprise 2 broad phases of processing (Klahr & Wallace, 1976; Trick & Pylyshyn, 1994; van Oeffelen & Vos, 1982). The first involves some kind of object individuation processing. Target items must be located, their features integrated and spatial locations determined. The second involves assigning a quantitative label, either to the entire set of objects at once, or iteratively to a subset of them, thus requiring subsequent summation. We will refer to this as the enumeration phase, although the term is used in an atheoretical sense and no particular enumeration strategy is implied. The first of these two phases must employ domain-independent processes because object identification or individuation is not a task that is specific to enumeration. However, the second phase is certainly enumeration-specific. This distinction has been expanded on by the research of Trick & Pylyshyn (1993, 1994), who have linked the first phase to early attentional processing. A number of theories have attempted to explain subitizing's slope and limit by appealing to one or other of these phases as the locus of the characteristic subitizing performance profile. There is not space here to review them in detail, but we shall briefly present the key aspects of three theories to contrast their accounts of subitizing.

Theories of Subitizing

The first detailed processing account of subitizing, developed by David Klahr (Klahr, 1973; Klahr & Wallace, 1976), had the primary goal of explaining the subitizing slope. The theory was stated in the form of a production system model in which rules "unpack" clusters of objects before the size of the group can be determined. The basic process is the creation of symbols in "Semantic Short Term Memory" (SSTM) from what might be referred to as "signals" in "Visual Short Term Memory" (VSTM), which contains the output of iconic processing. The quantitative symbols created in SSTM name a particular set size of objects that has been recognized in VSTM by quantity-specific productions. Being a serial system, the more objects there are in VSTM, the longer the recognition takes. Thus the cycle time of production firing, assumed by Klahr & Wallace to be 45 ms, accounts for the subitizing slope. Therefore, Klahr & Wallace's account of the subitizing slope is that it comes from processes that deal directly with the physically present stimuli represented in VSTM, "an extremely short-term memory ahead of STM in the visual chain" (1976, p.43). However, no data were collected that could be used to directly evaluate the proposed model.

Mandler & Shebo's (1982) primary goal was to explain the subitizing limit in terms of "canonical pattern" recognition. They claimed that, during childhood, we learn that collections of 1 to 3 objects fall into regular configurations: 1 is a singleton, 2 a line and 3 a triangle. Our recognition of these patterns results in a subitizing limit of 3 objects with a flat response profile. Explaining subitizing in terms of pattern-recognition means that Mandler & Shebo also load their account on perceptual processes. Since this account assumes that recognition of each canonical pattern happens as a single event, no slope is predicted by this account. The subitizing limit is predicted to be the limit of canonical patterns, i.e. 3 items. While Mandler & Shebo's theory does attempt to account for both the subitizing limit and slope, the theory is challenged by a lot of data. In particular, Akin & Chase (1978) found subitizing slopes and varying spans using block stimuli, many of which were similar to Mandler & Shebo's dot patterns.

The most recent account of subitizing is the FINST theory (Trick & Pylyshyn, 1993; 1994). This account primarily aims to explain the subitizing limit as a side-effect of the object individuation process in early vision. In this respect, the theory shares much in common with Klahr & Wallace's, since objects must be individuated before they can be enumerated. The main difference is that, in the FINST theory, object individuation takes place in parallel in the preattentive stage of vision. Here, object markers, or FINSTs, are attached to targets in the visual field for later processing, such as enumeration. Thus the subitizing slope prediction is exactly the reverse of Klahr & Wallace's. Since FINSTs are assigned in parallel, the slope can only arise from post stimulus-dependent processing. Trick & Pylyshyn (1994) state that "the response choice rather than the variable binding accounts for the subitizing slope" (p. 89). However, like Klahr & Wallace's account, there are no data that can be used to prove or disprove the operation of the FINST mechanism. The limit of subitizing is predicted to be the limit of the number of FINSTs that can be assigned. This is claimed to be 4 but no reason is given as to why it is not 5 or 3 or why it could not vary within or between individuals.

In this paper we present an analysis of subitizing performance that focuses on what we will call stimulus-limited processing. The label comes from Salthouse's (1981) definition of stimulus-limited and response-limited processes. Stimulus-limited processes are those which are assumed to operate when "the subject's processing [is] presumably controlled by the duration of the stimulus, and the subject [has] an unlimited time to respond" (Salthouse, 1981. p. 44). Response-limited processes are assumed to be those that operate when "stimuli [are] presented until the subject [makes a] response and it [is] the occurrence of the response that limited the information processing" (Salthouse, 1981. p. 44).

The methods that we have used to assess subitizing performance fit Salthouse's definition of tasks for measuring stimulus-limited processing, and we think that they provide insight into subitizing for the following reason. While the stimulus-limited/response-limited distinction does not necessarily imply a decomposition of subitizing into object individuation and enumeration phases, it seems reasonable to assume that object-individuation processes would be stimulus-limited in nature. This is because the quality of the representations that perceptual processes are able to create would be limited to information that could be extracted while the stimuli were available for inspection. Therefore, limited exposure, masked stimuli, like those we presented should enable us to examine the effects of increasing numerosity on stimulus-limited processes. If subitizing is based on object individuation processes, one would predict an increase in stimulus-limited processing with numerosity, as did Klahr & Wallace. However, if the phenomenon is due to response selection, as Trick & Pylyshyn propose, then increasing number should not affect stimulus-limited processing in a significant manner. Therefore, we believe that analyzing the stimulus-limited characteristics of subitizing will provide one way to determine which existing account offers the best prospects of progress towards a fully specified information processing account of subitizing.

Differential Time-Accuracy Functions

Our methodology of individual subject analysis based on Time Accuracy Functions (TAFs) has been described in detail elsewhere (Simon, Cabrera & Kliegl, 1993). However, a brief description is required here in order to explain our results.

Prior to each experiment we use pilot-testing to generate a *diagnostic range* of presentation times for each of the numerosities 2 through 8. Such a range comprises 10 presentation times where the shortest produces close to chance performance and the longest produces close to perfect enumeration. Nine of the times are separated by the same increment and a tenth "dummy" time is added with a larger increment. This provides maximal overlap with ranges for other numerosities, thereby reducing the chance that subjects can detect a numerosity on the basis of presentation time alone. From each subject's responses (based on 1400 trials) we generate a TAF showing the accuracy profile with respect to presentation time. In our earlier work (Simon et al., 1993), this relationship was then modeled using a negatively accelerated exponential function. Subitizing produces very steep functions, rising from chance to near-perfect detection with minimal additions of presentation time. Other enumeration strategies, used for larger numerosities, are reflected in shallower to almost linear functions. Figure 1 presents TAFs from a subject in the current experiment who subitized the numerosities 2 through 4.

The analysis that allows reaction time decomposition into stimulus-limited and post-stimulus components is described in detail by Cabrera & Simon (1995). In summary, individual subject TAFs are taken as input to analyses whereby Differential Time Accuracy Functions (DTAFs) are computed to produce measures of stimulus-limited processing. Each TAF is interpreted as the cumulative probability distribution of the time a subject needs to successfully complete a given task. By differentiating this distribution with respect to time, the corresponding probability density function can be computed. From this we can estimate, among other things, the mean (or mathematical expectation), and the standard deviation of process duration. The mean represents the maximum likelihood estimator for the duration of the stimulus-limited processes.

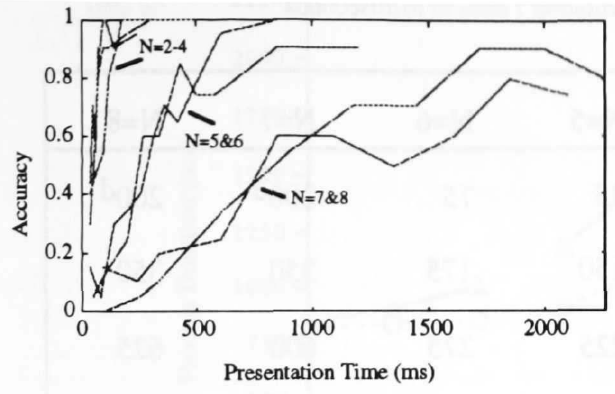

Figure 1: Time Accuracy Functions from Subject # 29

The standard deviation reflects the complexity of these processes. A small standard deviation is indicative of a unitary, simple process whereas a large standard deviation is taken to indicate the combination of several separate processes. We will refer to these measures as the subject's **mean duration** and **standard deviation of duration**. These measures do not account for the entire reaction time of the subject, but only for the time it takes to extract the perceptual information that is needed. The rest of the operations that occur correspond to what we call post-stimulus processes. We assume these to be independent of, and sequentially related to stimulus-limited processes. Therefore, subtracting mean duration from total response time gives an estimate of post-stimulus processing for the task in question. It is important to note that these post-stimulus processes are not the same as response-limited processes. As Salthouse (1981) makes clear, response-limited processes cannot be measured unless the subject has control over presentation time by virtue of terminating it with a response. This was not the case in the experiment reported here. Therefore, we define post-stimulus processing as the mental activity that takes place subsequent to stimulus-limited processing.

We are interested here in 2 main questions. First, are the slopes achieved by combining stimulus-limited and post-stimulus measures together similar to previous reaction time slopes? If they are, this would provide evidence that our analyses produce credible data for the study of subitizing. Second, what are the relative contributions of stimulus-limited and post-stimulus processes to the subitizing and post-subitizing slopes? In other words, we should be able to determine whether the subitizing slope is produced primarily by stimulus-limited processes, by post-stimulus processes or by some combination of both. We shall report data from an experiment where subjects were shown dots in linear arrays of varying numerosities for a range of presentation times. A number of stimulus configurations have been used in subitizing studies, all of which have their strengths and weaknesses. Our larger research project is examining the effects on processing of presenting a range of these. The linear arrays used in this experiment have the benefit of removing cues to number such as subjective grouping or familiar patterning, and they are similar to those used in previous experiments (Atkinson, Campbell & Francis, 1976; Atkinson, Francis & Campbell, 1976; Svenson & Sjoberg, 1983). Yet, they have the disadvantage that line-length may provide numerosity cues. However, a control study that we have carried out provided evidence that line-length alone is a very weak cue to absolute numerosity.

Method

Materials

All experiments were carried out using an Apple Macintosh IIfx computer with 13 inch high-resolution RGB display. Stimulus presentation and data collection were controlled by the Cedrus Superlab software package. The stimuli were 9-point helvetica type font letter Os arranged without typed spaces between them. They appeared as black circles on a white background centered horizontally and vertically on the screen. The longest row (N=8) was 5/8 inch in length. The mask that was employed was a row of 16 identical letter Os that was 1 1/8 inches length. Response time measures were collected by means of a voice-activated relay.

Subjects

Twenty-eight undergraduate students who enrolled in introductory psychology classes volunteered as subjects in return for extra credit and completed the procedure. There were 17 males and 11 females. All had normal or corrected vision.

Procedure

Instructions were presented to subjects on the computer and were read in a self-paced manner. An experimenter was on hand to answer any questions that arose at this point or after completing 10 trials sampled from the task itself. These were employed to show subjects what they could expect to experience during the experiment. Stimuli were presented at eye level, 30 inches from the subject and subtended less than 2 degrees of visual angle, to ensure foveal presentation. Each trial began with the presentation of a fixation arrow for 1250 milliseconds. The head of the arrow pointed to the location where the center of the stimulus to be presented would appear. At the offset of the stimulus arrow, a stimulus was presented for its predetermined presentation time and then was masked. Any response emitted by the subject was not recorded until the mask appeared. However, once the stimuli were masked, the timer was stopped by the subject's spoken the numerical response. At this point the subject typed the same number he or she had just spoken. This was stored by the Superlab program which used it as a signal to begin the next trial. Subjects were instructed to respond as quickly as possible without sacrificing accuracy. Trials were presented in blocks of 70 (7 numerosities each with 10 presentation times) with trials randomized within each block. At the end of every block the subject could rest and continue the experiment when s/he was ready. The program requested that the subject take at least a 2 minute break after blocks 8 and 16. The presentation times that provided diagnostic ranges for this experiment were determined during pilot testing and are shown in Table 1

Table 1: Diagnostic Ranges of Presentation Times in Milliseconds

Time	N=2	N=3	N=4	N=5	N=6	N=7	N=8
T1	35	35	35	75	75	200[d]	200[d]
T2	45	48	70	150	175	350	350
T3	55	61	105	225	275	600	625
T4	65	74	140	300	375	850	900
T5	75	87	175	375	475	1100	1175
T6	85	100	210	450	575	1350	1450
T7	95	113	245	525	675	1600	1725
T8	105	126	280	600	775	1850	2000
T9	115	139	315	675	875	2100	2275
T10	250[d]	300[d]	500[d]	850[d]	1100[d]	2350	2550
increment	10	13	35	75	100	250	275

[d] Dummy time

Results

Mean duration of stimulus-limited processes was calculated for each individual subject for all 7 numerosities (2-8). These were then subjected to a single factor (numerosity) repeated measures analysis of variance. Results showed a significant increase of mean duration F(6, 162) = 164.07 p<.001. This demonstrates that, as numerosity increased, the mean duration of stimulus-limited processing also increased. We then carried out regression analyses on stimulus-limited and post-stimulus processes. Post-stimulus processing times were computed based only on correct trials. In order to compare our findings to those of previous studies we carried out regressions independently on the numerosities 2-4 (to represent the subitizing range) and 5-8 (to represent the range in which counting and other processes are executed). Within the subitizing range the results were as follows. The slope for stimulus-limited processes was 59.7 ms per item, r = .6, F(1, 82) = 123.0 p<.001, while the regression for post-stimulus processes was not significant, r = .08, F(1, 82) = 0.55 n.s. This suggests that all of the impact of increasing numerosity within the subitizing range was found in stimulus-limited processes alone. Outside the subitizing range a different pattern of results was found. The slope for stimulus-limited processes was 227.6 ms per item, r = .6, F(1, 110) = 157.8 p<.001. The regression for post-stimulus processes was also significant, with a slope of 130.3 ms per item, r = .5, F(1, 110) = 37.2 p<.001. When stimulus-limited and post-stimulus processes are combined additively, as discussed by Cabrera & Simon (1995), we find a striking resemblance in our results to those from previous studies. Total processing showed slopes of 59.7 milliseconds per item in the subitizing range and 357.9 milliseconds per item outside the subitizing range. These times are precisely in the range found by traditional reaction time measures of subitizing. This suggests that our methodology, which controls presentation time and hence response latency, does not interfere with subject's response patterns, even for the rather long presentation times associated with the larger numerosities.

These results appear to provide 2 kinds of evidence that our methodology and analyses are useful tools for the study of subitizing. First, our estimates of overall reaction time agree with those from traditional reaction time studies showing slopes of 40 to 120 ms per item and 250 to 370 ms per item inside and outside the subitizing range

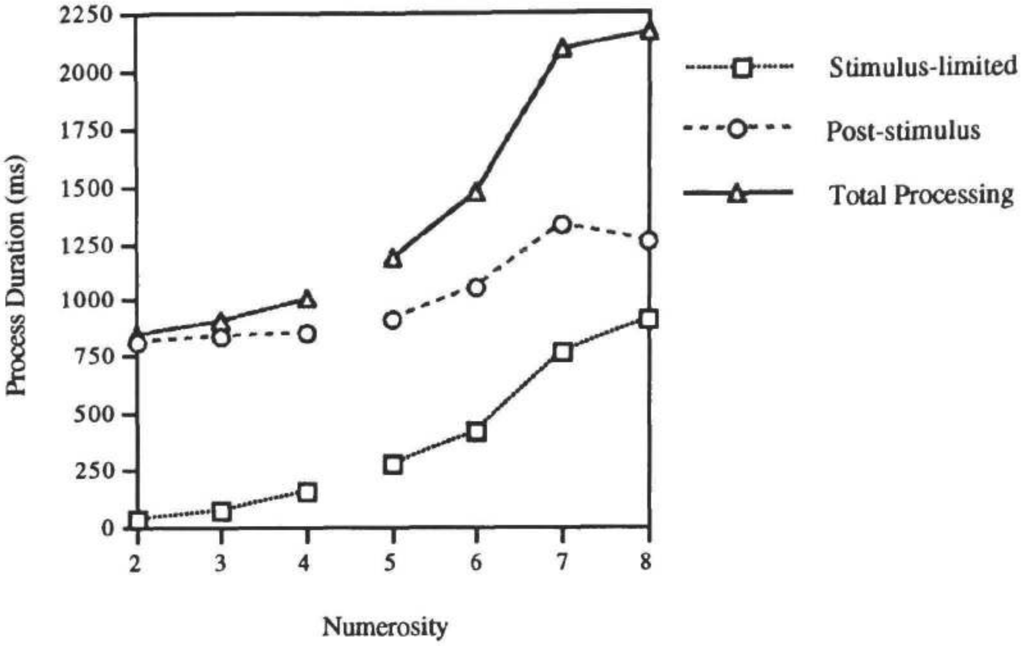

Figure 2: Total and Decomposed Process Times

respectively (Trick & Pylyshyn, 1994). Second, and more importantly, the results reveal which class of processes produce the subitizing slope. Figure 2 shows a plot of the duration of stimulus-limited processes, post-stimulus processes, and overall processing duration inside and outside the subitizing range. As can be seen, the subitizing slope is solely a phenomenon of stimulus-limited processes. There is a 59.7 millisecond slope for those processes in the subitizing range and no slope for the post-stimulus processes in the N = 2 - 4 range. These latter processes only begin to play a role outside the subitizing range, but in the current study only provide 36% of the overall slope.

Discussion

The experiment reported here produced results based on aggregated data that are very similar overall to those reported from previous studies. The slopes inside and outside the subitizing range produced by combining those of stimulus-limited and post-stimulus processes are exactly of the expected magnitudes. More importantly, the decomposed processes show that the subitizing slope is solely a function of increased stimulus-limited processing as numerosity increases in the subitizing range. Outside the subitizing range these processes still comprise 2/3 of the slope, but now post-stimulus processes play a role also. These results support the subitizing model proposed by Klahr & Wallace (1976), that increased numerosity generates the characteristic response time slope by affecting those processes that operate on the perceptual characteristics of the available stimuli. It is still unclear precisely what these processes are. However, our research is proceeding with a similar assumption to that of Trick & Pylyshyn; that object individuation processes are likely to be involved. Nevertheless, our results counter Trick & Pylyshyn's model, which suggests that response selection processes are responsible for the subitizing slope. Such selection processes would surely fall into the post-stimulus and not the stimulus-limited class. In other words, Trick and Pylyshyn's model would predict post-stimulus processes to show a slope in the subitizing range, rather than stimulus-limited processes. Figure 2 depicts exactly the opposite picture. The results also counter Mandler & Shebo's model, which predicts no slope at all within the subitizing range.

Furthermore, our results appear to reveal new piece of evidence. A typical finding in reaction time measurements of subitizing is that response latency for the largest numerosity often decreases, especially when subjects know what that numerosity is. The generally accepted explanation is that subjects employ a guessing strategy when many objects are presented. It is assumed that they do not attempt to quantify the larger arrays, but simply respond with the highest possible number. Hence the decision not to quantify the stimulus objects is reflected in lowered response time (as well as more errors). Our data show that, for the subjects we tested, this account is not strictly accurate. It can be seen from the Figure 2 that stimulus-limited processing increases with every numerosity increase, and very markedly for the largest numerosities. Thus, it appears that our subjects did try to quantify all the arrays that were presented, even the very largest ones. Presumably, they ran out of presentation time before object individuation could be completed for the array containing 8 dots and decided against investing resources in enumerating an array that they had failed to examine completely. Simply guessing the largest number is still likely to produce a correct response on some trials. This account is consistent with the dip in duration of post-stimulus processes for the last numerosity. Further studies which considerably lengthen presentation times for the largest numbers should provide data that would enable the evaluation of such an hypothesis.

We believe that the DTAF analysis we have presented is a very promising one for the investigation of the subitizing phenomenon. The results presented here are still rather preliminary, and further research is in progress on studies using other stimulus types. However, our data do suggest that subitizing linear arrays is entirely a function of increased stimulus-limited processing in response to increases in numerosity up to 4 items. In order to properly study the role of post-stimulus processes in these tasks, further research is needed where a self-terminated presentation time design is used. This would enable the examination of what Salthouse (1981) called response-limited processes. However, if our present results are found to be stable, then we already have a prediction for such studies. Response-limited processes are presumably not those primarily involved in object individuation and so would be predicted to only show effects of increased numerosity outside the subitizing range.

Acknowledgements

We would like to thank Christa Dell, Alan Kersten, Reinhold Kliegl, Jonah Lunken & Sandeep Vaishnavi for their input.

References

Akin, O. & Chase, W. (1978) Quantification of three-dimensional structures. *Journal of Experimental Psychology: Human Perception & Performance* **4**, 397-410.

Atkinson, J., Campbell, F.W. & Francis, M.R. (1976) The magic number 4 plus or minus 0: A new look at visual numerosity judgements. *Perception* **5**, 327-334.

Atkinson, J., Francis, M.R. & Campbell, F.W. (1976) The dependence of the visual numerosity limit on orientation, colour, and grouping in the stimulus. *Perception* **5**, 335-342.

Cabrera, A. & Simon, T.J. (1995) Time-accuracy data analysis: Separating stimulus-limited and post-stimulus processes. *Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Lawrence Erlbaum Associates.

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

Geary, D.C. (1995) Reflections of evolution and culture in children's cognition: Implications for mathematical development and instruction. *American Psychologist* **50**, 24-37.

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

Klahr, D. (1973) Quantification processes. In W. Chase (Ed.) *Visual Information Processing*. New York: Academic Press.

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

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

Salthouse, T.A. (1981) Converging evidence for information-processing stages: A comparative-influence stage-analysis method. *Acta Psychologica* **47**, 39-61.

Simon, T., Cabrera, A. & Kliegl, R. (1993) A new approach to the study of subitizing as distinct enumeration processing. *Proceedings of Fifteenth Annual Meeting of the Cognitive Science Society.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Simon, T.J. & Klahr, D. (1995) A computational theory of children's learning about number conservation. In T.J. Simon & G.S. Halford (Eds.) *Developing Cognitive Competence: New Approaches to Process Modeling*. Hillsdale, NJ: Lawrence Erlbaum & Associates.

Svenson, O. & Sjoberg, K. (1983) Speeds of subitizing and counting processes in different age groups. *Journal of Genetic Psychology* **142**, 203-211.

Trick, L.M. & Pylyshyn, Z.W. (1993) What enumeration studies can show us about spatial attention. Evidence for limited capacity preattentive processing. *Journal of Experimental Psychology: Human Perception & Performance* **19**, 331-351.

Trick, L.M. & Pylyshyn, Z.W. (1994) Why are small and large numbers enumerated differently? A limited-capacity preattentive stage in vision. *Psychological Review* **101**, 80-102.

van Oeffelen, M.P. & Vos, P.G. (1982) Configural effects on the enumeration of dots: Counting by groups. *Memory & Cognition* **10**, 396-404.