Modeling a Functional Explanation of the Subitizing Limit

Scott Peterson, Kimberly Morton, and **Tony J. Simon**
School of Psychology
Georgia Institute of Technology
Atlanta GA 30332-0170
psg95sp@prism.gatech.edu

Abstract

We present a model of enumeration that demonstrates one possible explanation for the limited capacity of subitizing. This analytical approach can be contrasted with most previous research on subitizing which has been primarily descriptive in nature, and which has tended to assume a structural limitation on the phenomenon. Our simulation results suggest instead that the limitation may arise from the functional constraints of learning to optimize among enumeration strategies for a space whose combinatorics increase greatly with number.

Introduction

The subitizing phenomenon has been a topic of interest and debate among psychologists for well over 100 years (e.g. Jevons, 1871). Kaufman, Lord, Reese, and Volkman (1949) coined the term subitizing to refer to the ability of adult human subjects to rapidly and accurately enumerate up to 3 or 4 discrete entities. The most typical characterization of subitizing is in terms of a shallow slope, of about 50 ms/item, in the aggregate reaction time data of subjects enumerating these small sets. This is contrasted with a far steeper 250 300 ms/item slope for enumeration of larger sets. The larger slope is taken to indicate the deployment of more complex processes such as counting. Therefore, the subitizing limit appears as a sharp discontinuity in reaction time measures at around 3 or 4 items (Atkinson, Campbell, & Francis, 1976; Chi & Klahr, 1975; Svenson & Sjoberg, 1983).

In recent years, researchers have become interested in developing a qualitative information processing account to explain the quantitative characterization described above (Dehaene & Cohen, 1994; Simon & Vaishnavi, 1996; Trick & Pylyshyn, 1993, 1994). As Simon & Cabrera (1995) point out, an adequate theory of subitizing must include an explanation for both the subitizing slope and the range of numerosities to which subitizing can be applied. There is now some consensus that the subitizing slope is in part due to rapid individuation of small, simple displays through processes associated with visual attention (Simon & Vaishnavi, 1996, Trick & Pylyshyn, 1993, 1994). However, there has been a notable absence of explanations of why the subitizing range is limited to 3 4 entities as opposed to 6 or even 2. Most attempts to explain the subitizing range thus far have appealed to *structural* limitations of the human information processing system. These are explanations typically involving some fixed capacity mechanism.

For example, Trick and Pylyshyn (1993) argued that subitizing is enabled by the assignment of a limited number of attentional tags, called FINSTs, to items in a visual display. They claim that since human adults have a maximum of 4 FINSTs, the number of items that can be subitized is also 4. Previous computational models of subitizing have also employed structural limitations. Klahr and Wallace's (1976) model contained productions specifically written to recognize collections of 1 through 3 items. Anderson, Matessa, and Douglass' (1995) ACT-R model of subitizing also used special productions to recognize small collections of one, two, and three items. An additional production allowed for one-by-one counting of items exceeding an initially recognized three. In that model, latencies were directly assigned to the three pattern recognition productions to produce the typical 50 ms slope within the subitizing range while a steeper slope was obtained outside the subitizing range due to firing of the item-by-item counting production. Thus, the subitizing limit as well as the slopes within and outside the subitizing range were pre-specified, or built-in to the model.

An alternative approach, which we present in this paper, is to explore the possibility that the subitizing range may reflect a *functional* rather than structural capacity limit. By a functional capacity limit we mean one that arises out of the interaction between processing characteristics attributed, without predetermined limits, to the agent, and the nature of the information being processed. In other words, an emergent rather than pre-specified property of the system.

To explore this functional hypothesis we have developed a computational model that simulates the emergence of the subitizing phenomenon as a result of learning to select optimal candidate enumeration strategies. As Siegler (e.g., Siegler & Shipley, 1995) has shown, children can optimize in this way by learning, not from the failure of candidate strategies, but by computing the relative accuracy and efficiency of each one with respect to given tasks. In a similar way, our model learns to select between two enumeration strategies depending on the numerosity it is presented with. The candidates we have implemented thus far will be referred to as *recognition* and *counting*. The recognition strategy roughly corresponds to subitizing and involves execution of a simple pattern-matching procedure that matches a newly presented pattern to a previously stored one whose numerosity is known. The counting strategy involves execution of item-by-item processing where, at each step, a unit within the pattern is "visited" and a running total is incremented. Through extensive training, our model learns to separate the problem space into two regions: One contains patterns that can be recognized, while the other contains patterns that must be counted.

An ACT-R Model of Enumeration

We developed our model using the ACT-R (Anderson, 1993) production system. ACT-R is a general theory of human cognition which assumes that cognitive processing is carried out through production rules operating on declarative memory. Our model takes advantage of two important features of the ACT-R system: conflict resolution and base level learning. Conflict resolution is a mechanism that determines which of a set of matched productions to select for execution. Candidate production instantiations are evaluated in terms of their expected values. The expected value of a production instantiation is, roughly, the value of the goal that can be achieved by firing the production minus the cost of firing it. The production instantiation with the highest expected value wins. Ties are resolved in favor of the production instantiation that can match its condition most quickly to declarative memory.

Learning in ACT-R can be accomplished in a variety of ways. For the present model we have chosen to focus solely on learning base level activations of the stored patterns in declarative memory. Base level activations are values associated with declarative memory elements indicating how "active" they are. In other words, base level activation is a measure of the strength of an item's memory trace based on recent processing. In general, higher activation implies faster retrieval. The effect of base level learning is to produce increases in activation of declarative memory elements as they are matched or retrieved. There is also a general decrease, or decay over time of these values. The rate at which this decay occurs is controlled by a global parameter in the ACT-R system which we have set to the value 0.1, representing a relatively low rate of decay. In combination with the conflict resolution mechanism described above, base level learning allows for a scenario in which strategy choice is mediated by activation levels of declarative memory elements. That is, one strategy applies when activation is below a certain threshold and another applies when activation is above that threshold. In our model, the recognition strategy requires that the base level activation of a stored pattern exceeds a certain threshold value. Once this threshold is reached, conflict resolution selects recognition over counting because the pattern-matching production can match more quickly to declarative memory than can the counting production. Repeated execution of the counting strategy serves to increase base level activations of the counted patterns to the point where recognition can take over. However, the combinatorics of the domain has the effect that only a subset of presented numerosities have patterns that can maintain the threshold level of activation over a period of time.

As a starting point, we have programmed our model to incorporate the enumeration knowledge of a 3- or 4-year old child who has two available enumeration strategies recognition, and counting (see Fuson, 1988 and Siegler, 1991 for reviews of the enumeration capabilities of preschoolers). The recognition strategy is modeled through a single production which matches a new pattern to a previously stored pattern from which the numerosity can be directly retrieved. The counting strategy is modeled by a small set of productions that sequentially visit unprocessed objects in a pattern and accumulate the total. Counting facts are provided to allow for sequential assignment of number names to objects. Each pattern represents one possible configuration of up to 6 objects on a 4 x 4 grid of locations.[1]

The model operates as follows. For each training pattern, the recognition and counting strategies participate in a competition to produce an enumeration. Early in training, activations on the stored patterns are low. This represents the assumption that children will have low confidence in their ability to recognize the numerical value of any given pattern until some learning has taken place. Thus, early in training, the counting strategy dominates, leading to predominant use of counting for all patterns.

Training serves to increase activations of the stored patterns. Whenever the counting strategy is applied to enumerate a training pattern, the final step is to increase the activation of the corresponding stored pattern in declarative memory. This represents increasing familiarity with one of a set of possible patterns for that numerosity. After a sufficient number of exposures to a particular training pattern, the corresponding stored pattern becomes active enough that the recognition strategy will win the strategy competition. Successful recognition, like successful counting also generates an increase of activation, or familiarity, for the enumerated pattern. The number of stored patterns in declarative memory for each numerosity is shown in Figure 1. This represents all possible patterns on a 4 x 4 grid.

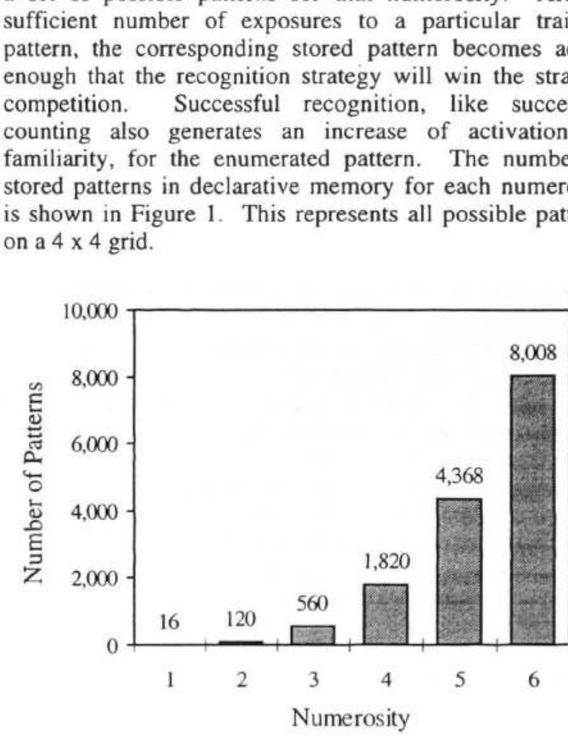

Figure 1: Number of stored patterns for each numerosity.

[1]The 16 cell grid was employed due to current limitations in computing resources, and is not theoretically motivated.

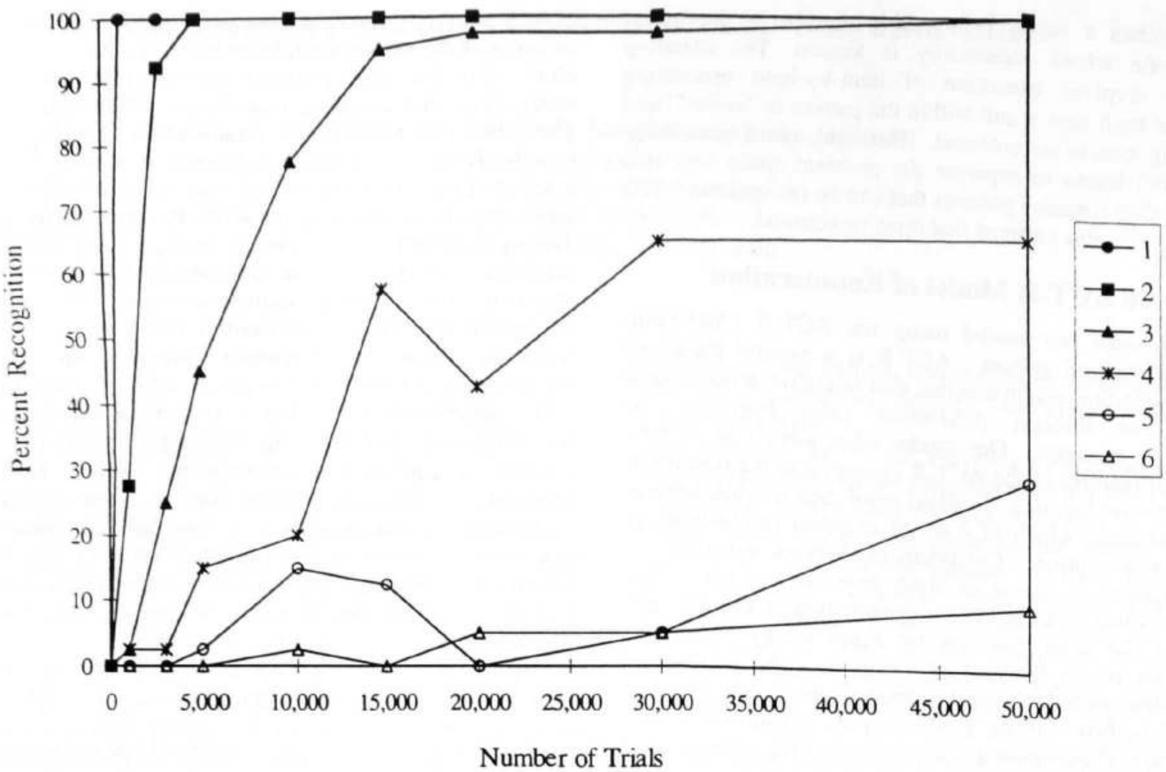

Figure 2: Percentage use of the recognition strategy for each numerosity as a function of the number of training trials.

With extensive training, the small numerosities (N = 1-3) as a whole become active enough such that the recognition strategy is generally used for all of their patterns. Large numerosities (N = 4-6), on the other hand, fail to reach this level of activation because of the relatively large number of possible patterns. For example, given our 4 x 4 grid for patterns, there are 560 possible patterns of 3 objects, but 1820 possible patterns of 4 objects. Thus, the number of training patterns required to achieve an overall increase in activation level for numerosity 4 is at least three times larger than the number required for numerosity 3. Furthermore, decay of activations becomes an increasingly important factor as the number of possible patterns grows. This is because the increasing number of patterns leads to a greater time delay between successive occurrences of each individual pattern. For many of the larger numerosity patterns, decay will decrease activation such that the recognition threshold is never reached. Thus, use of the counting strategy dominates for larger numbers. This behavior of the model after extensive training is consistent with empirical data suggesting that 5-year olds subitize small collections and count larger ones (Chi & Klahr, 1975).

The Simulation

Training and Testing

We conducted several simulation runs, each consisting of a training phase and a test phase. The length of the training phase was varied across the different simulation runs, from 200 to 50,000 training trials. The purpose of this training was to demonstrate the effects of increasing amounts of training on stored pattern activations, enumeration strategy choices, and enumeration latencies.

During the simulation, execution of each training trial proceeds as follows. First, a random numerosity between 1 and 6 is selected and a random pattern for that numerosity is generated. The pattern is then presented to the model to be enumerated by either the recognition or counting strategy. Either type of enumeration results in a strengthening of the activation for the stored pattern corresponding with the enumerated test pattern. Stronger activation of the stored pattern increases the likelihood that an identical test pattern will be recognized on some later trial. Currently, each test pattern must be presented at least twice before it will be recognized. More than two presentations may be required, however, depending on the time (number of trials) between presentations, since every stored pattern is subject to decaying activation during trials where it is not presented.

The last 120 trials of each modeling run constituted the test phase. During this phase, 20 random patterns for each of the six numerosities were presented to the model. Strategy choice and enumeration latency data (as computed by ACT-R) were collected during each test trial. After the completion of the test phase, the average activation level was computed for each numerosity.

Results

Figure 2 shows the effects of training on strategy selection. After very little training, patterns for numerosity 1 are enumerated using the recognition strategy exclusively. Patterns for numerosities 2 and 3 require somewhat longer training periods, but eventually are also enumerated exclusively by the recognition strategy. Patterns representing numerosities 5 and 6 continue to be primarily counted, even after thousands of training trials[2]

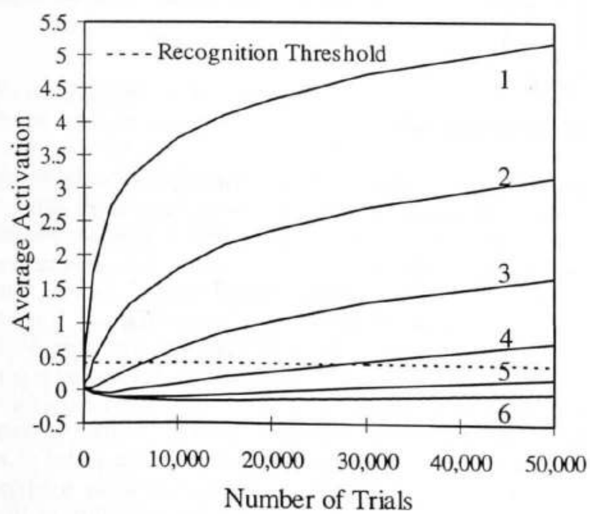

Figure 3: Average stored pattern activation for each numerosity as a function of the number of training trials.

The effect of training on the average activation level for the stored patterns of each numerosity is shown in Figure 3. All stored patterns begin with a base level activation of close to zero (i.e. 0.038). During training, activations of stored patterns for numerosities 1 3 quickly increase (on average) to a level above the threshold (depicted as a dashed line at 0.405 in Figure 3) required for exclusive use of the recognition strategy. Therefore, recognition becomes the primary strategy used for these numbers. The stored pattern activations for the larger numbers, however, do not reach a sufficiently high level for recognition to occur consistently. As described above, this is a result of the greater number of stored patterns for larger numerosities, and the decay which occurs because each individual pattern is presented less often.

For numerosities 5 and 6, counting continues to be the primary strategy, with recognition occurring intermittently and only for those relatively few patterns that happen by chance to be presented several times. Our simulation runs so far have indicated that the recognition rate for numerosity 4 appears to reach a maximum somewhere near 65%. The fact that this numerosity is neither conclusively counted nor recognized is consistent with experimental results. For example, Svenson & Sjoberg (1983) produced regressions for their 1 3 range and 5 8 range but were unsure with which range to associate $N = 4$.

Figure 4 shows the average latencies for a set of test trials (40 random patterns for each numerosity) before and after 50,000 training trials[3] Before training, the counting strategy is used exclusively, resulting in a linear increase in latency of about 1 second from one numerosity to the next throughout the entire range of numerosities. It takes nearly 3 seconds to enumerate one object and about 8 seconds to enumerate 6.

After training, recognition is used almost exclusively for numerosities 1 3, resulting in a small increase in latency as numerosity increases in this range. In contrast, a relatively large slope, similar in magnitude to the pre-training slope, is obtained for numerosities 4 6, reflecting extensive use of the counting strategy for these numerosities. The difference in slopes for the two ranges of numerosities is at least qualitatively consistent with experimental response time data. Improvement of the quantitative fit will be addressed in a future version of the model as discussed below.

Discussion and Conclusion

These results suggest a functional explanation for the origin of the subitizing limit. While most previous models of subitizing have simply assumed a limit of 3 or 4 (e.g. Klahr & Wallace, 1976; Trick & Pylyshyn, 1993, 1994) we have demonstrated one way that this limit might emerge as a function of the combinatorics of the space of patterns interacting with a simple learning and decay mechanism. Repeated counting of the same pattern serves to increase its activation to the extent that, at some point, that pattern is familiar enough for the child to confidently employ the recognition strategy. For small numerosities, the number of possible patterns is sufficiently small that all patterns get seen enough times to raise their activations above threshold. Since the number of possible patterns increases dramatically for larger numerosities (as shown in Figure 1), it is not surprising that even rather large amounts of training are not sufficient to raise activations above the recognition threshold. Furthermore, decay becomes a factor for the

[2] The values shown in this graph, as well as in Figure 4, represent averages of two simulation runs at each level of training. Each simulation run has some variability due to random generation of training patterns.

[3] The latency associated with each training trial is analogous to response time for an experimental trial. However, the actual latency values generated by the model are not intended to replicate experimental values since the pattern matching and counting processes have not been closely modeled.

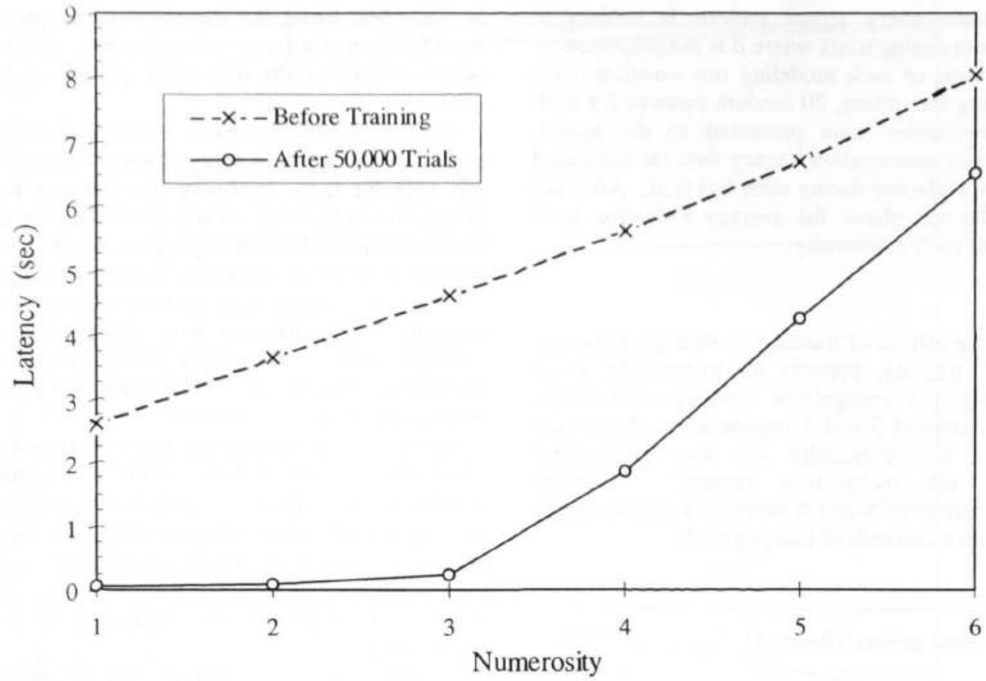

Figure 4: Average latency for each numerosity.

larger numerosities because more possible patterns also means more time between exposures to the same one.

In addition to this explanation for the limited subitizing range, our results also suggest a potential explanation for the origin of the subitizing slope. According to the ACT-R theory, response latency is composed of production matching latency and production action latency. Therefore, as matching latency increases, so does response time. Further, activations of declarative memory elements have a direct influence on production matching latency; the higher the activation, the shorter the matching latency. Thus, patterns with high activations will match faster than patterns with low activations, and therefore will result in shorter response latencies. This suggests that the subitizing slope may arise from differences in the average level of activation across numerosities.

Thus, we have been able to create a model which generates an explanation of the subitizing limit, and which may also be able to explain the slope. Unlike the previous models of Klahr and Wallace (1976) and Anderson, et al. (1995), our model is able to do this without pre-specifying either the subitizing limit or the subitizing slope. Rather, both emerge as by-products of learning along with the increases and decay in the activations associated with individual patterns.

Despite the promise of our results, we would like to mention some of the limitations of this model and suggest some possible future research directions. First, we have not attempted to address the mechanics of the recognition and counting processes in terms of detailed information processing specifications. Therefore, although we have qualitatively replicated the standard discontinuity in enumeration performance, the quantitative results produced by the model (such as the actual latency and slope values) do not match very well to empirical data at present. Future work will concentrate on trying to close this quantitative gap between simulation and empirical results. Second, this model does not address the issue of generalization. It must certainly be the case that children develop somewhat abstract representations of patterns so that recognition is not necessarily based on an exact match to a previously seen pattern. Perhaps matching a pattern to an abstract pattern representation is a third strategy that could be added to a future version of the model. Another possible solution might be to utilize the partial matching capability of the ACT-R system to allow inexact matching which is sensitive to the similarities between patterns. A further finding to be simulated, which would be a good test for the model, is the transition from counting to recognition of consistently presented large displays. There is clear evidence (e.g. Mandler & Shebo, 1982; Wolters, van Kempen, & Wilhuizen, 1987) that repeated presentations enable subjects to move from deliberate enumeration of individual items to recognition of entire patterns. We believe that the present model should be able to simulate that result. The principles upon which it would do so are consistent with automaticity theories based on consistent versus varied mapping (Shiffrin & Schneider, 1977) or the retrieval of familiar instances (Logan, 1988). In a future version of the model we would also like to examine the effect of increasing our assumed pattern grid size, e.g., from 4 x 4 to 6 x 6. Our expectation is that the same qualitative behavior patterns will emerge, although more training will be required due to the increased number of possible patterns for each numerosity.

Finally, an intriguing issue with respect to individual differences arises from our model's "indecisiveness" about which strategy to use for the numerosity 4. In its present incarnation the asymptote recognition level for 4 items is around 65%. To some extent, this probably reflects the fact that some larger patterns are just more "memorable" than others. However, the result is more interesting because not all human subjects subitize the same number of items, and there is no strict consensus on what the absolute limit of subitizing is, if indeed there is one. Since our model represents a functional rather than a structural explanation of the limit, we may be able to use other functional parameters to explain individual differences. For example, working memory span and information processing speed have both been implicated as functional limitations on processing. Varying approximations of these in future versions of the model should enable us to investigate the true nature of subitizing by modeling individual data, rather than the often misleading aggregated results which may obscure individual variance and characterize no single individual at all.

Acknowledgments

We are indebted to Noel Rappin for his participation in the design and construction of the original version of this model, and for the Lisp code used to generate and present the patterns, and to tabulate the output. We would also like to thank the ACT-R group at Carnegie Mellon, especially Christian Lebiere, for their assistance.

References

Anderson, J. R. (1993). *Rules of the Mind.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Anderson, J. R., Matessa, M., & Douglass, S. (1995). The ACT-R theory and visual attention. *Proceedings of Seventeenth Annual Meeting of the Cognitive Science Society.* Hillsdale, NJ: Lawrence Erlbaum Associates.

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.

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

Dehaene, S., & Cohen, L. (1994). Dissociable mechanisms of subitizing and counting: Neuropsychological evidence from simultanagnosic patients. *Journal of Experimental Psychology: Human Perception and Performance, 20*, 958-975.

Fuson, K. C. (1988). *Children's counting and number concepts.* New York: Springer-Verlag.

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

Kaufman, E., Lord, M., Reese, T., & Volkman, J. (1949). The discrimination of visual number. *American Journal of Psychology, 62*, 498-525.

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

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review, 95*, 492-527.

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

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review, 84*, 127-190.

Siegler, R. S. (1991). *Children's thinking.* Englewood Cliffs, NJ: Prentice-Hall.

Siegler, R. S., & Shipley, C. (1995). In T. J. Simon & G. S. Halford (Eds.), Developing cognitive competence: New approaches to process modeling. Hillsdale, NJ: Lawrence Erlbaum Associates.

Simon, T., & Cabrera, A. (1995). Evidence for subitizing as a stimulus-limited processing phenomenon. *Proceedings of Seventeenth Annual Meeting of the Cognitive Science Society.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Simon, T. J., & Vaishnavi, S. (1996). Subitizing and counting depend on different attentional mechanisms: Evidence from visual enumeration in afterimages. *Perception & Psychophysics, 58*, 915-926.

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.

Wolters, G., van Kempen, H., & Wilhuizen, G. (1987). Quantification of small numbers of dots: Subitizing or pattern recognition? *American Journal of Psychology, 100*, 225-237.