

Is subitizing a unique numerical ability?

J. D. BALAKRISHNAN

Northwestern University, Evanston, Illinois

and

F. GREGORY ASHBY

University of California, Santa Barbara, California

Two models that predict the relation between mean enumeration time and numerosity in a speeded enumeration experiment are tested. The first is a bilinear two-process model, and the second is a log-linear single-process model. Previously, support for the bilinear model has provided evidence for the existence of a unique numerical ability called "subitizing." Both models are shown to yield close approximations to the empirical data, but at the same time to consistently violate the robust shape of these data. Two fundamental discrepancies exist: (1) Enumeration of single-element displays is unpredictably fast, both in the data reported here and elsewhere, and (2) the response-time function for multiple elements is continuously convex upward, with a significant log-quadratic component. The findings support the contention that enumeration is a capacity-limited process, but no statistically reliable change in processing character, that is, from subitizing to some other process, is evident in enumeration of displays of up to six elements.

In the "speeded enumeration" experiment, subjects report the number of elements in a display as quickly as possible. Typically the duration of the display is brief (e.g., 200 msec), but the elements are of high contrast and easily discriminable from each other and the surrounding field. The classical results are that enumeration time is almost invariant for display sizes up to about four elements, but thereafter, it increases sharply as display size increases (Chi & Klahr, 1975; Kaufman, Lord, Reese, & Volkman, 1949; Mandler & Shebo, 1982). When display duration is limited, the response time (RT) versus numerosity function reaches asymptote at about eight elements, although error rate continues to increase (Kaufman et al., 1949; Mandler & Shebo, 1982). Three separable regions of the RT data are, therefore: (1) fast and slowly increasing from one to three or four elements (region 1), (2) moderately fast and rapidly increasing from four to six or seven elements (region 2), and (3) slow and unchanging at eight or more elements (region 3). This article investigates the numerical processes underlying regions 1 and 2.

The speeded enumeration paradigm has a long history in psychology. The fact that subjects enumerate four elements almost as quickly as one has been interpreted as support for (1) span of attention or "immediate apprehension" (Bourdon, 1908; Cattell, 1886; Jevons, 1871; Oyama, Kikuchi, & Ichihara, 1981); (2) a unique numer-

ical ability called "subitizing" (Chi & Klahr, 1975; Kaufman et al., 1949; Klahr & Wallace, 1976); (3) canonical pattern recognition (Mandler & Shebo, 1982; Neisser, 1967; Oyama, 1986); or (4) some form of visual perceptual discrimination (Saltzman & Garner, 1948; van Oeffelen & Vos, 1982). The relatively sharp increase in RT in region 2 has sometimes been included in one of these four categories, but it is more likely to be attributed to a distinct process called "incremental counting" (Mandler & Shebo, 1982). Region 3 is generally considered to reflect a gross estimation of "numerousness" (Kaufman et al., 1949) that depends on perceptual properties such as local density, rather than enumeration in a strict sense.

Much of the empirical support for each of these viewpoints rests on the exact relation between mean response time (MRT) and stimulus numerosity and also the degree to which this relation depends on experimental conditions (Folk, Egeth, & Kwak, 1988; Mandler & Shebo, 1982). Two quantitative and testable models of this relation have been proposed. The first postulates a bilinear dependence of MRT on numerosity (Klahr & Wallace, 1976), and the second postulates a log-linear dependence (Kaufman et al., 1949). Despite its theoretical significance, methods of evaluating the empirical MRT versus numerosity function have not been rigorous, in the sense that neither of these models has been subjected to statistical goodness-of-fit testing, nor has a direct comparison of the two approaches been attempted.

In this article, we test the validity of these two models with a rigorous statistical analysis and we show that neither provides a satisfactory account of the data. The results are then shown to be consistent with a large body of previous data, including both classical and more re-

This research was supported in part by National Science Foundation Grant BNS88-19403 to the second author. The first author was also supported by Grants MH44640 to Roger Ratcliff and AFOSR 90-0246 (jointly funded by NSF) to Gail McCoon. Correspondence concerning this article should be addressed to J. D. Balakrishnan, Department of Psychology, Northwestern University, Evanston, IL 60208.

cent studies. Finally, implications of the exact nature of the models' failures for theories of subitizing and for potential adjustments to the models are discussed.

THE BILINEAR MODEL

The first model postulates that MRT increases linearly with a small slope in region 1 and increases linearly with a significantly larger slope in region 2 (Chi & Klahr, 1975; Klahr & Wallace, 1976; Mandler & Shebo, 1982; Oyama et al., 1981). Formally, this model predicts that

$$\begin{aligned} RT_{i,j} &= \alpha_1 i + \beta_1 + n_{i,j}, & i \leq m, \\ RT_{i,j} &= \alpha_2 i + \beta_2 + n_{i,j}, & i > m, \end{aligned} \quad (1)$$

where i is the stimulus numerosity, j is the trial number, m is 3 or 4, $n_{i,j}$ is a random variable with a mean of zero, and α_k and β_k are free parameters. Note that the noise is indexed by trial and stimulus numerosity to account for the fact that the empirical RT variance increases with increasing numerosity.

The change in slope between the two functions in the Equation 1 model is seen as *prima facie* evidence for a capacity-limited process that operates on the smallest displays (Chi & Klahr, 1975; Klahr & Wallace, 1976). The term "subitizing," originally chosen by Kaufman et al. (1949) to describe enumeration of up to six elements, was adopted by Klahr and his colleagues to emphasize the distinction between the two linear regions. This usage has become widely accepted.

The shallow linear slope of the function in region 1 is often thought to reflect the operation of a specialized numerical ability, perhaps present even at birth (Starkey & Cooper, 1980; Starkey, Spelke, & Gelman, 1983). However, an alternative view is that subjects learn to recognize canonical geometric patterns existing in two-dimensional displays of small numerosity (one element is a "dot," two elements form a "line," and so forth), and then use these geometric properties to enumerate the displays quickly (Mandler & Shebo, 1982; Neisser, 1967; Woodworth & Schlosberg, 1954). In the present paper, the stimuli did not include canonical display patterns, but enumeration was nevertheless rapid and accurate for displays of up to about four elements.

THE LOG-LINEAR MODEL

The second quantitative model proposes that there is no discontinuity between regions 1 and 2, but instead that enumeration time is a continuous, exponential function of numerosity. Formally,

$$RT_{i,j} = \beta e^{\alpha i + n_{i,j}}, \quad (2)$$

where $1 \leq i \leq 7$, β and α are free parameters, and $n_{i,j}$ is defined as before (Kaufman et al., 1949; von Szeliski, 1924).¹ Equation 2 is an example of a log-linear model because after the natural log of both sides is taken, it becomes

$$\log(RT_{i,j}) = \alpha i + \log(\beta) + \eta_{i,j}.$$

Thus $\log(RT)$ is a linear function of numerosity.

Various physical properties of the stimulus might suggest the Equation 2 model; for example, the relative change in stimulus energy with numerosity. If the model holds, then it could be construed as evidence that enumeration processes are dominated (with respect to the temporal demands of the task) by perceptual processes and that there is no differentiated subitizing mechanism.

The results reported below suggest two fundamental problems with the quantitative models just described. First, enumeration of single-element displays is uniquely fast, leading to a significant discontinuity in MRT between one and three elements that is not predicted by either model. Second, although the MRT function at numerosities 2-6 is continuous and convex upward, ruling out the bilinear model, it has a sharper bend than the exponential family of functions can predict (i.e., it has at least a log-quadratic component and possibly higher components as well). This complexity, we believe, should discourage attempts to use the RT data as evidence for a distinct numerical ability.

To test the validity of these two models, a speeded enumeration task was run in which the stimuli were horizontally linear arrays of solid-colored rectangular blocks. This arrangement eliminates geometric cues to numerosity for all except the single-element displays (which could be seen as "dots" instead of "lines"), and thus permits a test of the canonical-pattern-recognition hypothesis (see above). In one condition, the elements were equally spaced and the total length of the display varied. In the second condition, length was fixed and the distance between elements varied. To reduce the correlation between physical properties of the display, such as total energy or contrast in the image, from numerosity, the colors of the individual blocks were also varied. (Note that local density would also have some correlation with numerosity in the fixed-length condition. As it turns out, conclusions to be drawn below are not dependent on differences in the two display conditions.)

METHOD

Subjects

Eighteen subjects participated. Eight subjects were current or former student employees at the University of California, Santa Barbara, and 10 were undergraduate students satisfying part of an introductory psychology course requirement.

Stimuli

The stimuli were horizontally linear arrays of one to seven solid-colored blocks presented on a grayish-white background. The display monitor had a vertical refresh rate of 60 Hz. Seven colors were used: red, green, blue, yellow, pink, brown, and gray. Luminance values were 7.30, 21.89, 56.62, 28.31, 62.46, 4.90, and 21.30 cd/m² for the seven colors, respectively. The stimuli were selected randomly for each trial, without replacement. The blocks were 1.0 cm in width and 2.7 cm in height and were presented at approximate eye level at a viewing distance of approximately

200 cm. Visual angle of the individual blocks was 0.29° horizontally and 0.77° vertically.

Two display-pattern conditions defined the location and separation of the blocks: *fixed interval* and *fixed length*. In the fixed-interval condition, the distance between blocks was fixed while end-to-end distance varied. In the fixed-length condition, end-to-end distance was fixed while distance between adjacent blocks varied. Both conditions were based on an implied template of nine possible locations in a linear array, separated by 1.4 cm from edge to edge.

In the fixed-interval condition, locations were chosen from left to right sequentially, with the starting position chosen randomly and constrained by the total number to be displayed. The total visual angle subtended by the array varied from 0.29° to 5.78° ; the interval between blocks was a constant 0.40° .

In the fixed-length condition, when more than one element was presented, the left- and right-most locations from the template were always included and the remaining blocks were located randomly at any of the intervening seven positions. End-to-end visual angle was a constant 5.78° , and intervals between blocks varied from 0.40° to 5.21° . For one-element displays, the position was chosen randomly from the nine possible locations. Note that in this condition, the single-element display has a more central location on average than the two-element display, but this is not true for the fixed-interval condition.

Procedure

This experiment was performed concurrently with a second experiment (not reported here), with the order counterbalanced across subjects and the two sessions separated by 1 full week. The subjects wore a lapel microphone, which was connected to a voice key with an input to the controlling computer. Speed of response was emphasized, and the subjects were informed that the maximum number of elements presented would be seven. To begin each trial, the subjects depressed the space bar on the keyboard in front of them. They were instructed to enter at this keyboard whatever response they had actually vocalized for each trial, regardless of whether or not they had subsequently changed their minds.

The experiment included 14 stimulus conditions (7 numerosities \times 2 display patterns), and display duration was fixed at a constant 200 msec. The displays were presented in a random order for 70 practice trials (5 per condition), followed by 280 test trials (20 per condition). A short break was allowed at the end of the 70 practice trials and after 140 test trials. Total session time was approximately 45 min.

RESULTS AND DISCUSSION

Percent correct for the data combined over subjects is shown in the first (top) panel of Figure 1. Panels 2 and 3 present MRT and RT standard deviation, respectively, for correct responses in the combined data.

Note that enumeration times are relatively fast and accurate for numerosities up to four elements for both display conditions; at higher numerosities, efficiency drops off sharply. To this extent, the data are similar to previous reports (Atkinson, Cambell, & Francis, 1976; Chi & Klahr, 1975; Kaufman et al., 1949; Mandler & Shebo, 1982). Further corroboration of this assessment will be provided by the fitting analyses described below.

Within-subject analysis of variance on RT by numerosity and display condition was performed to investigate the effects of display condition. This showed significant

effects of numerosity [$F(6,2154) = 410.85, p < .001$], display condition [$F(1,359) = 39.31, p < .001$], and their interaction [$F(6,2154) = 4.639, p < .001$]. Because the single-element displays are the same in both display conditions, data from these trials were removed and the remaining data were partitioned according to whether numerosity was small (2-4) or large (4-7). As suggested in Figure 1, analysis of this partitioned data showed that the numerosity \times display condition interaction is due to performance on smaller numerosities [interaction values: $F(2,718) = 8.96, p < .001$, for numerosities 2-4; and $F(3,1077) = 1.79, p > .148$ for numerosities 4-7]. Visual angle is larger in the fixed-length than in the fixed-interval condition, and the difference decreases with numerosity, which might account for some of the differences between the two conditions. Nevertheless, the basic shape of the two functions is similar; that is, they both appear to be convex upward. It is this shape, rather than the quantitative differences, that will determine the validity of the enumeration models.

Figure 1 also suggests unique effects of the stimulus numerosities 1 and 7 (the extremes of the numerosity dimension). Specifically, the slope of the function decreases: $MRT_2 - MRT_1 > MRT_3 - MRT_2$, and $MRT_6 - MRT_5 > 0 > MRT_7 - MRT_6$. Both results are fairly consistent across subjects, the first being violated in only 8 of 36 cases (18 subjects \times 2 display conditions) and the second in only 5 of 36 cases. Various strategic behaviors could be postulated to account for these results. Perhaps the simplest explanation is that numerosities 1 and 7 have only one immediately adjacent response alternative (e.g., numerosity 1 has neighbor 2), whereas all other numerosities have two. Thus, any information that favors an impossible response (e.g., 0 or 8) could be taken as strong evidence for numerosity 1 or 7. The single-element data could also be explained by the difference in display pattern associated with single-element versus multiplelement displays (single elements could be seen as "dots" as opposed to "lines"). Both accounts illustrate how learned abilities that do not reflect strictly numerical skills might significantly affect enumeration performance. Because neither of the two statistical models described above is able to predict these effects, further analyses focused on numerosities 2-6.

Consider first the Equation 1 model discussed above. For several reasons, the most appropriate test of this model would assume that the subitizing limit is four elements (i.e., $m = 4$ in Equation 1). Any other hypothesis would appear to strongly violate the observed data in Figure 1. Further, since at least three data points are necessary to test the assumption of linearity, by choosing a subitizing limit at numerosity 4, it is possible to determine whether any two contiguous regions in the empirical functions satisfy the bilinear model. If two such regions exist, then at least one of the two estimated functions under this model must be consistent with the data.

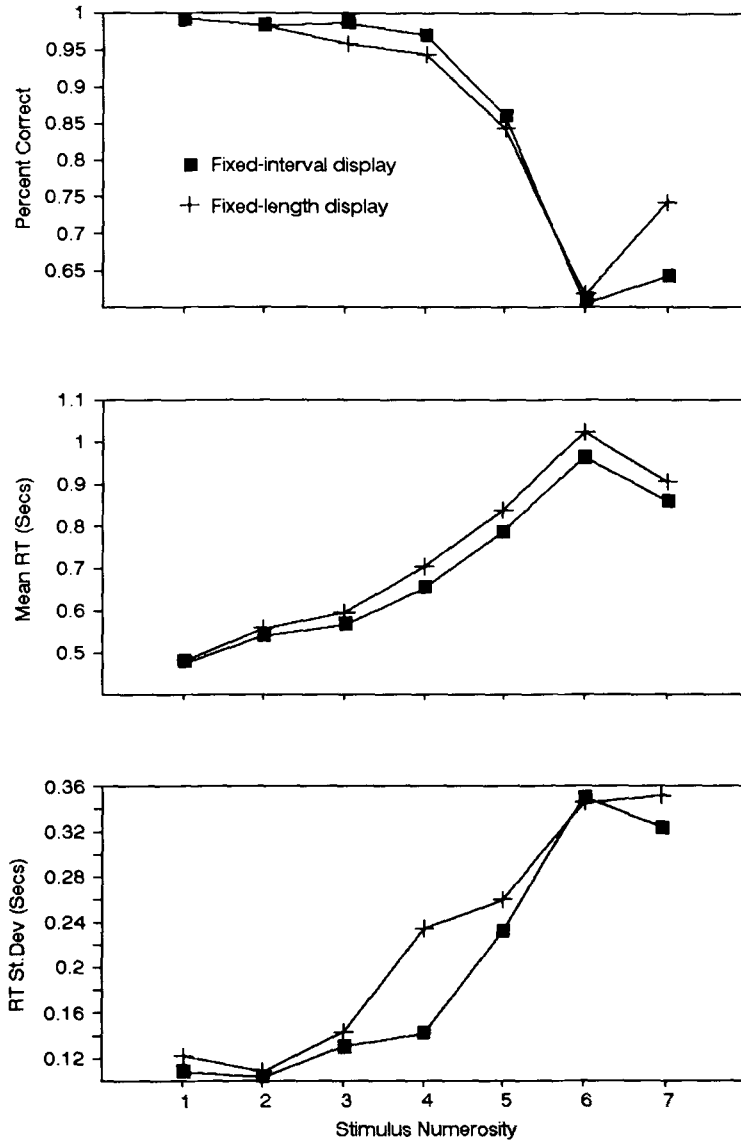


Figure 1. Summary statistics for speeded enumeration of linear displays of 200-msec duration. The response-time measures are based on correct responses only; minimum $N = 218$.

To optimize the parameter estimates for this version of the bilinear model, the fits were obtained using a weighted least squares algorithm that minimizes

$$\begin{aligned}
 WSSE = & \sum_{i=2}^4 \sum_{j=1}^N \left[\frac{RT_{ij} - (\alpha_1 i + \beta_1)}{\sigma_i} \right]^2 \\
 & + \sum_{i=4}^6 \sum_{j=1}^N \left[\frac{RT_{ij} - (\alpha_2 i + \beta_2)}{\sigma_i} \right]^2, \quad (3)
 \end{aligned}$$

where σ_i is the RT standard deviation when the numerosity is i (to be estimated from the data). Note that Equation 3 takes into account the increase in RT variance as-

sumed in the model by putting less weight on those data points with greater estimation error.

The upper panel of Figure 2 shows the best-fitting linear functions for the Figure 1 data at numerosities 2-4 and 4-6. To a first approximation, the model seems quite adequate. However, a closer inspection reveals a tendency for the empirical values to be underestimated at the endpoints and overestimated in the middle. Note that this is precisely what is to be expected if the entire function from 2-6 is continuous and concave upward.

To verify the consistency of this result across subjects, the best-fitting functions were also computed for each subject individually. The bottom panel of Figure 2 shows the averaged residuals (i.e., errors of prediction) at each data

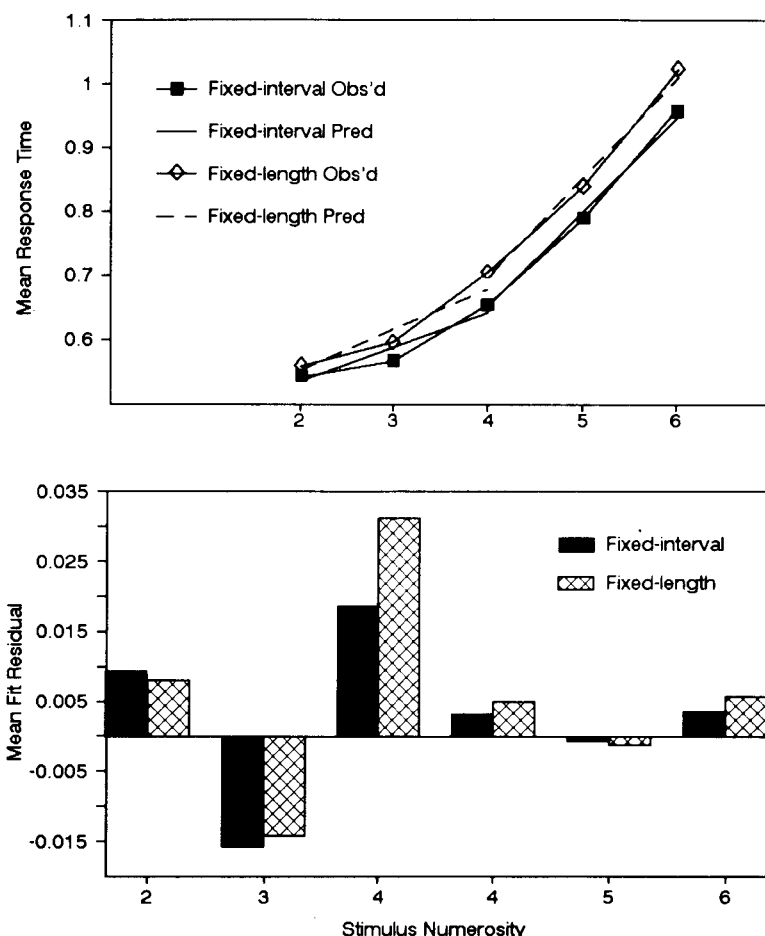


Figure 2. Weighted least squares estimates of a bilinear model for mean response time (same data as in Figure 1). The lower panel shows the mean prediction error (observed minus predicted) over fits to individual subjects' data for each data point.

point. (Note that this method is different from fitting to the combined data set because of the weighting function). The results are similar in form and even more robust in comparison to the previous test. However, they are also misleading in one case. Specifically, for displays of fixed length at numerosities 4-6, 12 of the 18 subjects show the opposite result from that suggested in Figure 2 (i.e., overestimation of the endpoints and underestimation in the middle). This result compares with only 2 deviations from the Figure 2 pattern in the fixed-spacing condition for numerosities 4-6, and 7 out of 36 for the two other functions combined. Figure 3 illustrates these results with four examples from individual subjects.

If the bilinear model is correct, then the errors of prediction (i.e., the residuals) should be normally distributed with a mean of zero. This hypothesis provides another means of statistically corroborating the failures of the bilinear model. Kolmogorov-Smirnov tests on the normality of the residuals collected over the 18 subjects were therefore performed for each of the four linear fits. Departures from normality were significant in each of the four

cases (maximum $p = .014$), providing further evidence that although the violations of the bilinear model are small, they have a consistent form.

These results suggest that the log-linear model might fare better, since it predicts a continuous MRT versus numerosity function. However, the log-linear model performed no better than the bilinear model. Weighted least squares estimates on the observed log(RT) data are shown in Figure 4. Note that the data strongly suggest at least a quadratic trend and possibly higher components in log(RT). The effect is also consistent across subjects and display conditions. The lower panel of Figure 4 shows the averaged residuals for each data point. Only 5 of the 36 total tests differ noticeably from the pattern shown in Figure 4.

To facilitate comparison with the bilinear model, the data were divided into the same four groups as before (numerosities 2-4 and 4-6 \times two display conditions) and Kolmogorov-Smirnov tests for normality were performed on best fits to these. Interestingly, despite clear violations evident in Figure 4, these tests were nonsignificant for

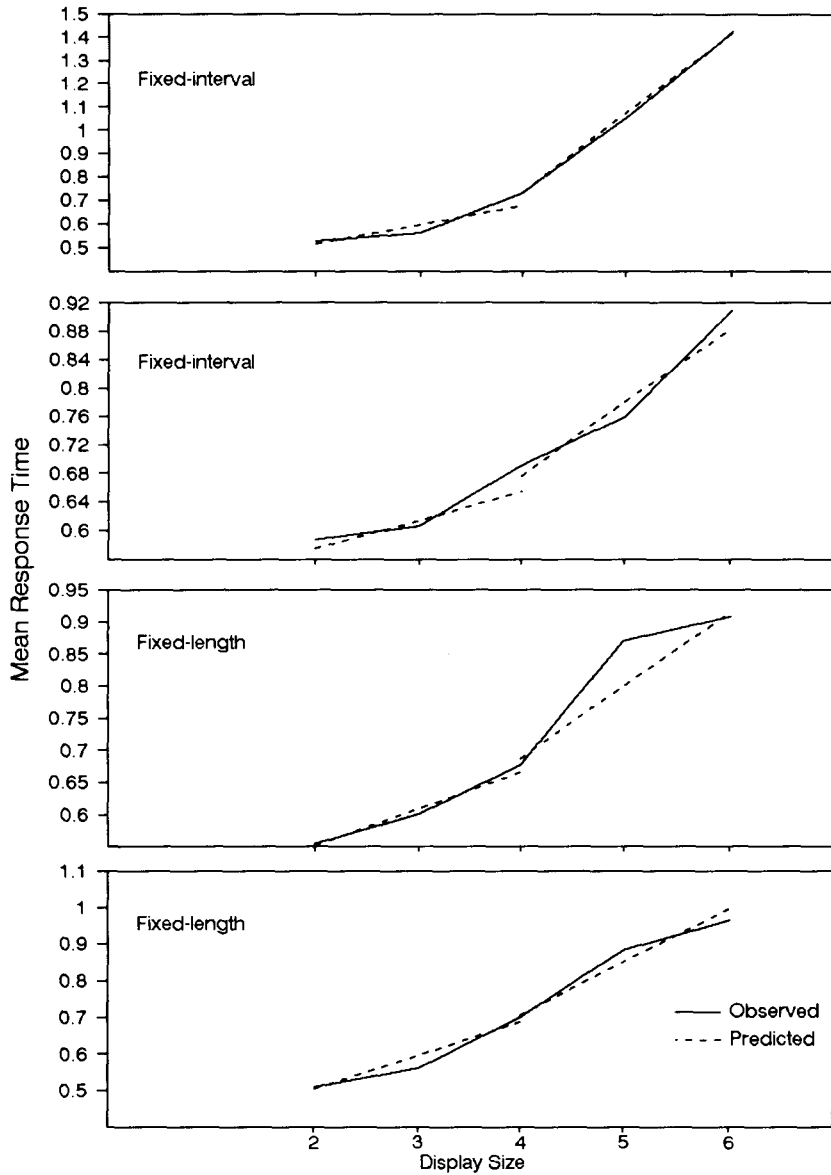


Figure 3. Representative examples of the bilinear model fit to individual subject data.

numerosities 4-6 ($p > .08$ for the fixed-interval and $p > .55$ for the fixed-length displays). Significance was reached, however, at the smaller numerosities ($p < .01$), and so the log-linear model is nevertheless strongly rejected.

In addition to the models described in Equations 1 and 2, three related models were also tested. The first model assumed that enumeration time is a power function of numerosity:

$$RT_{i,j} = \beta i^\alpha + n_{i,j}, \quad (4)$$

for some constants α and β . The second and third models were versions of the log-linear and power-function mod-

els, with one additional parameter. The three-parameter log-linear model assumed

$$RT_{i,j} = \beta e^{\alpha i + n_{i,j}} + \gamma, \quad (5)$$

where α , β , and γ are free parameters. A third free parameter was added in a similar way to the power-function model of Equation 4. Although we omit the details, very similar conclusions were obtained for each of these models. Specifically, the predictions of each model deviated from the data in a manner similar to the log-linear model. Thus, each of these models can be rejected.

Both the bilinear and log-linear models, as well as several related ones, are seen to be unequivocally rejected

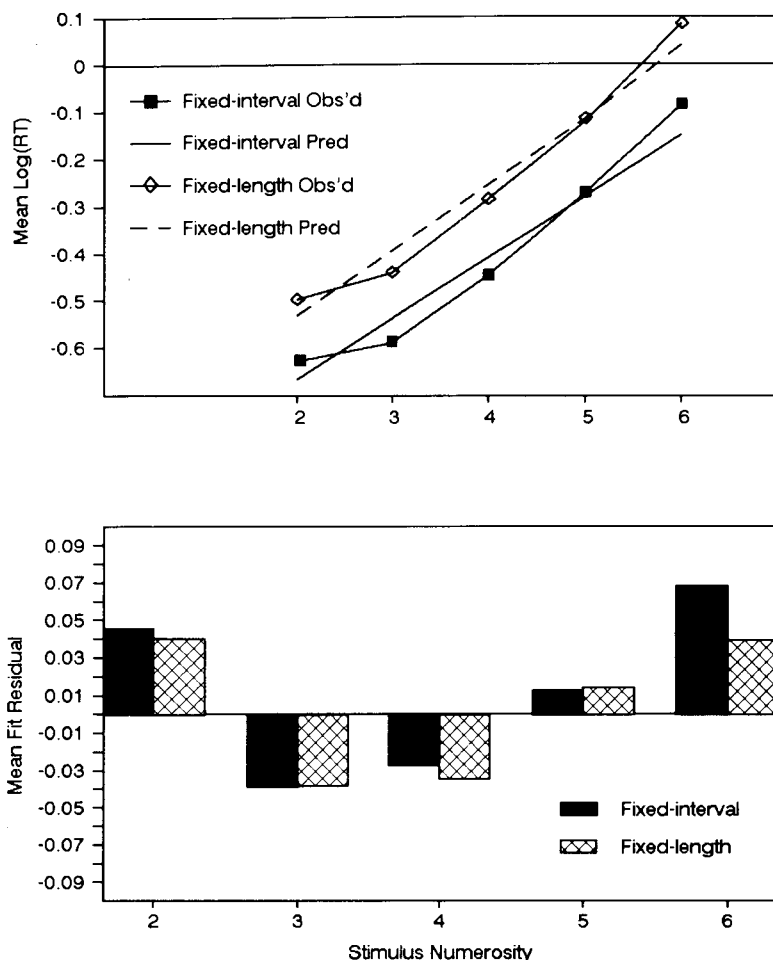


Figure 4. Weighted least squares estimates of the exponential model for mean response time (same data as in Figure 1, and lower panel defined as in Figure 2).

by the present data. Before considering the implications about subitizing on the basis of these results alone, however, it is important to first address the issue of their generality. Of particular concern is whether the failure of these quantitative models could be due to an artifact of the experimental conditions of this particular study.

Such an explanation would seem unlikely for several reasons. First, the most important features that have previously been identified with subitizing (fast and accurate enumeration of small numerosities followed by sharp increases in RT and error rate) are also characteristic of the data in Figure 1. Therefore, if different mechanisms were invoked by the stimulus conditions of this study, then more than one kind of subitizing would be implied, even though both mechanisms must be highly efficient.

Second, these same properties appear to be robust with respect to a wide variety of experimental manipulations. For example, several studies have found no effect of numerosity range on what defines subitizing and when it is operating (Klahr & Wallace, 1976; Mandler & Shebo, 1982). The overall pattern of results also seems to de-

pend little on the developmental level of the subjects. Although their responses are slower overall, young children produce MRT versus numerosity curves of the same type as Figure 1 (e.g., Svenson & Sjöberg, 1978). Finally, Oyama et al. (1981) found apparent support for the bilinear model using displays of different durations followed by a pattern masker. Specifically, the authors proposed a two-stage model in which elements may or may not be encoded, but if they are encoded, then they are enumerated and RT increases according to the bilinear model.

There is little evidence, then, that the use or availability of a subitizing mechanism is particularly susceptible to variations in the stimulus conditions. The present data constitute further evidence against such a position because the fast and accurate enumeration of small displays was possible in the absence of geometric cues. The question that remains is therefore whether the exact nature of the model failures demonstrated above is also robust across changes in experimental design. If similar sorts of inconsistencies can be shown to exist in previous studies as well,

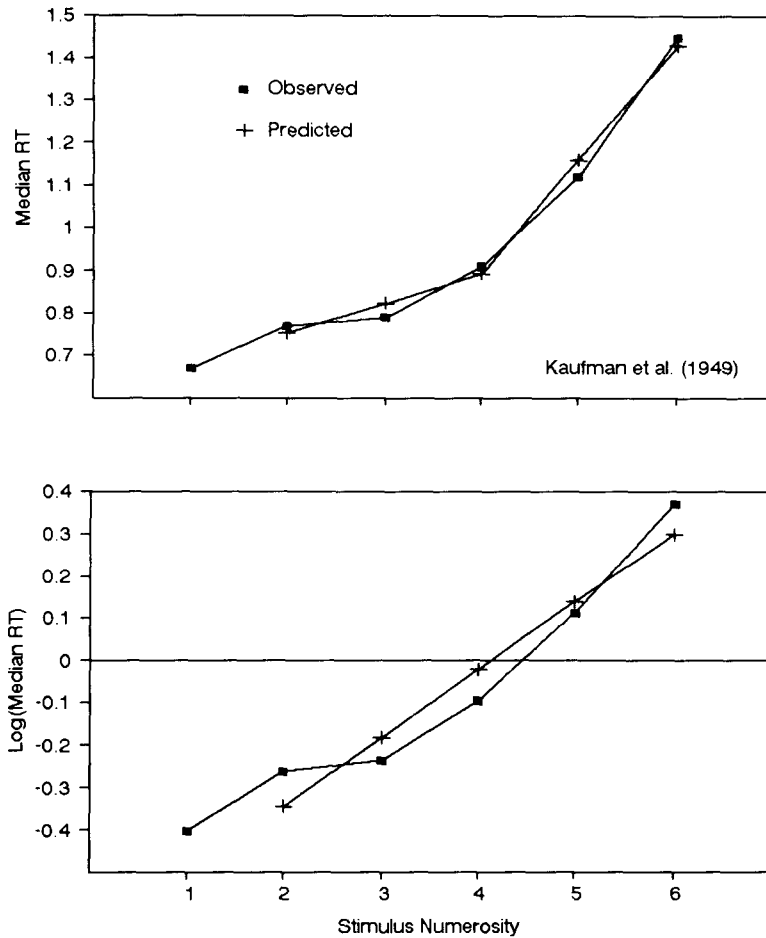


Figure 5. Best fitting bilinear and exponential models for data of Kaufman et al. (1949).

then this would constitute strong evidence that the bilinear and log-linear models must be rejected or revised.

The essential issue is the slope of the MRT function. If it increases continuously, then the bilinear model is ruled out, and if it increases, but not in a constant proportion to MRT, then the log-linear model is ruled out. Figure 5 shows the best-fitting functions to the often-cited data of Kaufman et al. (1949; from the condition in which response speed was emphasized). Each of the two fundamental problems reported above surfaces once again in these data: uniquely fast responses to single-element displays, and increasing but nonexponential slopes from 2–6 elements. Similar results are obtained if one applies the log-reciprocal transform that Kaufman et al. (1949) used. Note that these data were obtained from an experiment with a larger range of numerosities (1 to 15 elements) and the standard two-dimensional random-dot displays.

Other frequently cited data were reported by Jensen, Jensen, and Reese (1950) and by Mandler and Shebo (1982). The data of Jensen et al. (1950) were obtained using response-terminated displays, whereas Mandler and Shebo (1982) used a fixed display duration. In addition,

Mandler and Shebo (1982) investigated the effects of numerosity range in a series of three conditions (i.e., ranges 1–15, 1–6, and 1–3). Estimates of the MRT increase between each successive pair of numerosities are given in Table 1. Also included in the table are the two instruction conditions from Kaufman et al. (1949). All three data sets noticeably violate the bilinear predictions because of monotonically increasing slopes in the 2–6 range. The log-linear model also shows the same pattern of violations as before in all three data sets. What emerges from this survey, then, is not a consistent body of evidence in favor of one model or the other, but additional evidence contrary to both.

Implications

The function relating MRT and numerosity appears to be continuous and convex upward, with an acceleration that is too abrupt to be explained by any of the classical functions. The bilinear and the log-linear models each captured one important feature of the data, however. The bilinear model predicted the distinctively fast responses at small numerosities, and the log-linear model predicted

Table 1
Slope of Mean Response Time by Numerosity in Data of (1) Kaufman et al. (1949) (Median RT), (2) Jensen et al. (1950) (Median RT), and (3) Mandler and Shebo (1982) (Mean RT)*

N	(1)		(2)	(3)		
	Instructions		Response-Terminated	Range		
	Accuracy	Speed	Displays	1-15	1-6	1-3
1-2	90	100	50	30	-10	15
2-3	20	20	50	80	70	35
3-4	30	120	160	190	200	
4-5	380	210	260	300	220	
5-6	1,090	330	380	310	180	

*Estimates for Mandler and Shebo (1982) are based on their Figures 7 and 8.

a continuous increase in slope. Before any strong claims can be made about a special numerical ability, it is necessary to reconcile these two results.

Perhaps the simplest revision of the bilinear model that would maintain a dual process account and still predict a continuous function would be to assume that the capacity of subitizing is not fixed, but instead varies from subject to subject, or else from trial to trial for an individual subject. The latter hypothesis is more consistent with the results presented above, since the bilinear model also failed when it was fit to the data of individual subjects.

In this revised model, the MRT at each numerosity is a mixture of a subitizing and an enumeration process. Let $P(S|i)$ be the probability that a display of size i is subitized, and let $P(C|i) = 1 - P(S|i)$ be the probability that it is "counted." In addition, let $MRT_{s,i}$ be the mean RT on subitizing trials when the display size is i and let $MRT_{c,i}$ be the analogous mean RT on counting trials. These two values are the expected values of the first and second linear functions in Equation 1, respectively. Overall mean RT at numerosity i is therefore equal to

$$MRT_i = P(S|i) MRT_{s,i} + P(C|i) MRT_{c,i}. \quad (6)$$

With the additional free parameters, this model can account for the continuity in the averaged data, despite the fact that the component processes are linear.

Apart from the fact that it is untestable because of the added parameters, this mixture model has an interpretation problem. Specifically, the slope of the best-fitting linear function in the putative counting region is steeper than in the subitizing region (see Figure 2). Thus, throughout the subitizing range, the predicted counting time is less than the predicted subitizing time! One consequence of this fact is that the slope of the counting function cannot represent the time to enumerate an element in the display, nor can the counting function be ascribed to a single process that is distinct from subitizing. Otherwise, subitizing would be the *slower* of the two processes.

One way to avoid this difficulty is to assume that the quasilinear regions do not provide good estimates of the two slopes because of the change in subitizing capacity between trials. To fit the model, one could constrain each function to predict the same intercept, with this value being

the estimated time for a motoric response and other non-numerical processes. Although the constrained and unconstrained versions are both untestable with the extra mixing parameters of Equation 6, one way to examine these models in more detail is to determine the least squares estimates of the slopes of the two functions, under the constraint that they be as close as possible to the estimates from the Equation 1 nonmixture model (i.e., minimizing the amount of mixing predicted). This can be achieved by adding to the weighted sum of squared errors (WSSE) of Equation 3 the value $\alpha \max[P(S|i), 1 - P(S_i)]$, where α is a positive constant.

Table 2 shows the results of this analysis for each of the two mixture models just described. Note that the apparent capacity of the subitizing process differs significantly in the two models. In the single-intercept model it is about four or five elements, and in the two-intercept model it is three or four elements (depending on the criterion). Note also that the estimated slopes in the two-intercept model are comparable to the estimates in the nonmixture model, whereas the estimates for the single-intercept model are quite different. This difference suggests that the usual method of fitting the Equation 1 model

Table 2
Estimated Slope (α), y-Intercept (β), and Probability of Subitizing Given Numerosity i (P_i) for the Standard Bilinear and Bilinear Mixture Models Described in the Text

	Bilinear Mixture					
	No Mixture		Version 1*		Version 2†	
	Fixed Interval	Fixed Length	Fixed Interval	Fixed Length	Fixed Interval	Fixed Length
α_i	53	62	25	37	25	37
β_i	429	430	492	485	492	485
α_{ii}	148	156	155	159	79	90
β_{ii}	59	76	37	69	492	485
P_2			1.00	1.00	1.00	1.00
P_3			1.00	1.00	1.00	1.00
P_4			0.01	0.55	0.99	0.96
P_5			0.00	0.02	0.86	0.74
P_6			0.00	0.00	0.01	0.04

Note—"1" and "2" indicate regions of the linear fit (see text); slope and intercept values are in milliseconds. *Two intercepts. †One intercept.

provides little information about the capacity and speed of subitizing.

CONCLUSIONS

Each of the two existing models of speeded enumeration fails to adequately describe the dependence of enumeration time on stimulus numerosity. Because of its implicit assumption of two processes, the bilinear model constituted a major source of support for the existence of a unique numerical process in humans. Rejection of this model therefore significantly weakens the evidence for such a process. Failure of the log-linear models, on the other hand, rules out the class of single-process models that relate the temporal demands of enumeration to the proportional change in numerical magnitude as integer magnitude increases. The latter result is important because for larger numerosities than those used in the present study, proportional changes in integer magnitude appear to be quite good predictors of numerical judgments (van Oeffelen & Vos, 1982).

Although it is not unreasonable to suppose that some variance in the capacity of subitizing might account for the lack of a discontinuity in the MRT versus numerosity functions, there seems to be little reason to favor this position over more parsimonious accounts of the data, in which subitizing is seen as one of many manifestations of a general capacity limitation. In keeping with this view, Broadbent (1975) has compiled a long list of experiments in which at most three or four items can be processed with high efficiency in memory. Even granting that the RT data differ appreciably in some of these experiments, it is unlikely that separate structural mechanisms are necessary to account for each of them. The challenge raised by the present results is therefore to provide a more convincing demonstration that subitizing deserves its current status as a unique numerical mechanism.

REFERENCES

- ATKINSON, J., CAMPBELL, F. W., & FRANCIS, M. R. (1976). The magic number 4 ± 0 : A new look at visual numerosity judgments. *Perception*, *5*, 327-334.
- BOURDON, B. (1908). Sur les temps nécessaire pour nommer les nombres. *Revue Philosophique*, *65*, 426-431.
- BROADBENT, D. E. (1975). The magic number after fifteen years. In A. Wilkes & A. Kennedy (Eds.), *Studies in long term memory* (pp. 3-18). New York: Wiley.

- CATTELL, J. M. (1886). Über die Trägheit der Netzhaut und des Sehzentrums. *Philosophische Studien*, *3*, 94-127.
- CHI, M. T. H., & KLAHR, D. (1975). Span and rate of apprehension in children and adults. *Journal of Experimental Child Psychology*, *19*, 434-439.
- FOLK, C. L., EGETH, H., & KWAK, H. (1988). Subitizing: Direct apprehension or serial processing? *Perception & Psychophysics*, *44*, 313-320.
- JENSEN, E. P., JENSEN, E. M., & REESE, T. W. (1950). The subitizing and counting of visually presented fields of dots. *Journal of Psychology*, *30*, 363-392.
- 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*, *62*, 498-525.
- KLAHR, D., & WALLACE, J. G. (1976). *Cognitive development: An information processing view*. Hillsdale, NJ: Erlbaum.
- MANDLER, G., & SHEBO, B. J. (1982). Subitizing: An analysis of its component processes. *Journal of Experimental Psychology: General*, *111*, 1-22.
- NEISSER, U. (1967). *Cognitive psychology*. Englewood Cliffs, NJ: Prentice-Hall.
- OYAMA, T. (1986). The effect of stimulus organization on numerosity discrimination. *Japanese Psychological Research*, *28*, 77-86.
- OYAMA, T., KIKUCHI, T., & ICHIHARA, S. (1981). Span of attention, backward masking, and reaction time. *Perception & Psychophysics*, *29*, 106-112.
- SALTZMAN, I. J., & GARNER, W. R. (1948). Reaction time as a measure of span of attention. *Journal of Psychology*, *25*, 227-241.
- STARKEY, P., & COOPER, R. G. (1980). Perception of number by human infants. *Science*, *210*, 1033-1035.
- STARKEY, P., SPELKE, E. S., & GELMAN, R. (1983). Detection of intermodal numerical correspondences by human infants. *Science*, *222*, 179-181.
- SVENSON, O., & SJÖBERG, K. (1978). Subitizing and counting processes in young children. *Scandinavian Journal of Psychology*, *19*, 247-250.
- VAN OEFFELEN, M. P., & VOS, P. G. (1982). A probabilistic model for the discrimination of visual number. *Perception & Psychophysics*, *32*, 163-170.
- VON SZELISKI, V. (1924). Relation between the quantity perceived and the time of perception. *Journal of Experimental Psychology*, *7*, 135-147.
- WOODWORTH, R. S., & SCHLOSBERG, H. (1954). *Experimental psychology*. New York: Holt.

NOTE

1. Kaufman et al. (1949) actually fit a slightly different function to median response time; however, the differences are inconsequential.

(Manuscript received November 20, 1990;
revision accepted for publication August 3, 1991.)