The second number-estimation elbow:

Are visual numbers greater than 20 evaluated differently?

Makayla Portley and Frank H. Durgin

Swarthmore College

In press, January 28, 2019, in *Attention, Perception & Psychophysics*

Abstract

Numerosity perception has long been understood to be divided between subitizing and estimation. In a series of three experiments (total N = 113), a new number "elbow" point in the estimation of visual number for numerosities of about 20 dots is confirmed. Below 20, mean estimates are linear with a slope of about 1 and power-function exponents for numerosity estimation approximate unity, though estimate variance increases dramatically above about 6 elements. For numerosities above 20, estimates become increasingly compressed, such that power function exponents are much lower (e.g., 0.7) and are lower still when both ranges are estimated within the same experimental procedure. The experiments described here show that the location of the inflection point appears insensitive to the range of numbers estimated and to differences in density.

Author Note

Makayla Portley and Frank H. Durgin, Department of Psychology, Swarthmore College. Correspondence concerning this article should be addressed to Frank H. Durgin, Department of Psychology, Swarthmore College, 500 College Avenue, Swarthmore, PA 19081. E-mail: fdurgin1@swarthmore.edu

For over 150 years, it has generally been argued that the perception of non-symbolic visual number may be divided into two different regimes with the division occuring between low numbers (up to about 4 or 5 units), where number can be accurately and immediately perceived, and higher numbers where the uncertainty of estimation occurs (Atkinson, Campbell & Francis 1976; Fernberger, 1921; Hamilton, 1859; Jevons, 1871; Kaufman, Lloyd, Volk & Reese, 1949; Mandler & Shebo, 1982; Taves, 1941; Trick & Pylyshyn, 1994). For example, Kaufman et al. found that response latencies for the estimation of very small numbers, were brief; they increased rapidly as number increased beyond 4 items, and they plateaued at between 6 and 10 items. Similarly, as early as Jevon's experiment with cast beans 150 years ago, it appeared that the break from perfect accuracy in estimation occurred between 4 and 5.

Here we consider a second discontinuity in the estimation of number, a break (or elbow) in the number estimation function that seems to occur at about 20 items (Durgin, 2016). Whereas much recent work on perceived number has measured comparisons, rather than the enumerations, of visual collections (e.g., Burr & Ross, 2009; Dakin, Tibber, Greenwood & Morgan, 2011; Durgin, 1995; Gebuis & Reynvoet, 2012; Leibovich, Katzin, Harel & Henik, 2017; Van Oeffelen & Vos, 1982), the present investigation extends research traditions that have been primarily concerned with explicit number estimation (i.e., explicitly evaluating a multitude of elements) rather than visual comparison of quantities based on their apparent numeric magnitude. Although we will speculate about the meaning of this second elbow in the General Discussion, the main purpose of this paper is to characterize this second elbow by means of careful empirical investigation of relevant parameters, such as dot density and the range of values employed, that might affect it.

Kaufman et al. (1949) labeled the first "elbow" in numerosity perception (at about 5 or 6 dots) as the boundary between "subitizing" and "estimation", based on differences in response time, confidence, and accuracy (subitizing, from the Latin for sudden, was described as fast, accurate and associated with great confidence; estimation was usually not). Others have proposed that subitizing may be limited to only 4 elements, replaced by rapid counting up to six or seven (Atkinson et al., 1976), and this is consistent with Kaufman et al.'s response latency data. Although the "second elbow", an inflection in the number estimation function at about 20 dots, can (with hindsight) be discerned in the estimation data of Kaufman et al., (i.e., increasing underestimation as numerosity increased) they made no comment on it.

Krueger's (1972) study of numeric estimation for numbers 25 or greater showed that a power function with an exponent (log-log slope) of about 0.75 captured his data well. Krueger found only small effects of display area, though these might be interpreted as reflecting imperfect integration of density information and area. A power function with an exponent less than 1 implies increasing underestimation in this range, though it is not clear what the source of compression is. An exponent of 0.5 could be expected if, for example, observers substituted mean linear separation between elements for density; it remains possible that visuals systems seek to integrate a variety of sources of relevant information (density, area, linear separation, etc.; Gebuis & Reynvoet, 2012) when estimating numbers in this higher range. An alternative source of response compression, could be anchoring effects, when taking repeated measures. To rule out effects of repeated measures, Krueger (1982) had each of 800 participants make only a single numerosity estimate. He still observed an exponent less than one (i.e., 0.85). Krueger (1982; Figure 1) estimated that his power function would cross the unity line at about 20 dots, but he did not report direct investigations of the transition point.

More recently, in the course of an investigation of the effects of adaptation on perceived number, Huk and Durgin (1996) collected verbal estimates of visual number of homogenous elements both in adapted and in unadapted regions of the visual field (these fields of small white dots were centered 4° to each side of fixation). They confirmed that estimates were proportionally lower in the densely-adapted field (an effect that gradually emerged between 5 and 20 dots), and that estimates in each field produced elbow-shaped estimation functions in log-log space that seemed to sharply bend at about 20 dots (see Durgin, 2016). Whereas estimation data for numerosities up to 20 dots could be fit with a power function with an exponent somewhat greater than 1, the estimation data from 40 to

1152 dots, though offset in log-log space, as a function of adaptation, were well fit by power functions having an exponent of about 0.65. This exponent might be lower than observed by Krueger (1972) for many reasons, including response compression, given the large range of numbers tested within-subjects. Moreover, Huk and Durgin (1996; Durgin, 2016) did not vary the display area of their estimation fields, and their fields were presented briefly and peripherally, unlike those of Krueger. A departure from linearity in estimates is what characterizes the second number elbow observed by Huk and Durgin. It suggests a shift in the estimation process. Whether this shift is related to cognitive or to visual processes of estimation (or to some combination of the two) remains to be seen.

In the studies to be reported here we sought to measure this second number estimation elbow with centrally presented collections, and to evaluate whether the location of the elbow depended on density, number, or other factors. For example, Anobile, Cicchini and Burr (2014; see also Anobile, Turi, Cicchini & Burr 2015) have suggested that numerosity evaluation processes may vary as a function of density. Although density has been implicated in estimation of number (e.g., Dakin et al., 2011), our results will not tend to support the specific idea that a critical density defines the transition point marked by the second number elbow.

Many labs have chosen to create number displays with equal numbers of dark and light dots to avoid confounds with luminance (e.g., Burr & Ross, 2008), but this approach could easily affect number estimation by defining two equal subsets of dots for separate estimation. To avoid this concern, we instead used identical luminance-balanced elements (luminance-balanced dots; Carlson, Moeller & Anderson, 1984; Durgin & Huk, 1997) randomly positioned without overlap in one of three sizes of display so as to dissociate both density and brightness from number. Our participants were asked to estimate, as exactly as they could, the number of dots presented in each display; the displays were presented briefly to prevent counting.

Experiment 1: The second elbow in the estimation of visual number

In addition to seeking to replicate the second elbow, this experiment was designed to test whether the range of stimuli presented affects the scaling of number estimation above 20. That is, using the same kind of stimuli in both paradigms, does presenting numbers only above 20 (like Krueger, 1972) produce different kinds of estimates in the upper range than a design that includes lower numbers (like Huk & Durgin, 1996, and Kaufman et al. 1949).

To directly study the effect of range while discouraging a dependence on density alone and to rule out trivial issues of resolution, we adopted a design to meet these several constraints: (1) the annular elements had a bright center and dark surround that were of the same average luminance as the background gray, but were resolvable well into the periphery, so that all elements were identical, (2) the area in which the elements appeared varied over a range greater than 2 so that number would be dissociated from density, and (3) half the participants saw a quasi-logarithmic, but complete range of numerosites from 1 to over 200, similar to Kaufman et al. (1949), while the other half only saw numerosities greater than 25, similar to Krueger (1972).

Method

The procedures in this and the following experiments were approved by the local Institutional Review Board.

Participants. Forty-one undergraduate students (at least 18 years old and of both sexes) participated in partial fulfillment of a research requirement for the Introductory Psychology course. The data from one participant was excluded from analysis because the numerosity estimates given were extreme outliers (in some cases 20 SD above the means of the other participants), leaving 40 participants. Of these, half made estimates over the full-range; half only saw numerosities greater than 25.

Stimuli. The number displays were viewed from a distance of 60 cm on a 19" Sony CRT flat screen (34.3 cm display width) with a resolution of 1280 x 1024 pixels refreshed at 100 Hz. The experiment was controlled using Psychtoolbox (Brainard & Vision, 1997; Kleiner, Brainard & Pelli, 2007). Each element in the presented array was a white disc 0.30° (12 pixels) in diameter, surrounded by a black annulus, with an outer diameter of 0.46° (18 pixels). The background gray of the remainder of the screen was selected so as to match the average luminance of the elements as measured by a spot photometer (72 cd/m^2). Independent of their density, the dots for each display were randomly scattered based on an algorithm which required that the center of each new randomly-selected dot was at least 0.61° (24 pixels) from the centers of all other dots in the display, as well as from the central fixation point, so as to avoid overlap and minimize clustering. Finally, the distance of each dot from the center of the display could be no more that that required to ensure that the outer edge of the dot was within the proscribed radius of the presentation area, which was either 7.6°, 9.5°, or 11.4° (300 pixels, 375 pixels or 450 pixels). Sample displays are shown in Figure 1. A single black cross hair at the center of the screen was used to facilitate fixation prior to the onset of the display. Once all dot positions had been computed, the dot array was displayed for 400 ms and then replaced again by the fixation display with a text box for entering the estimate using a keyboard.

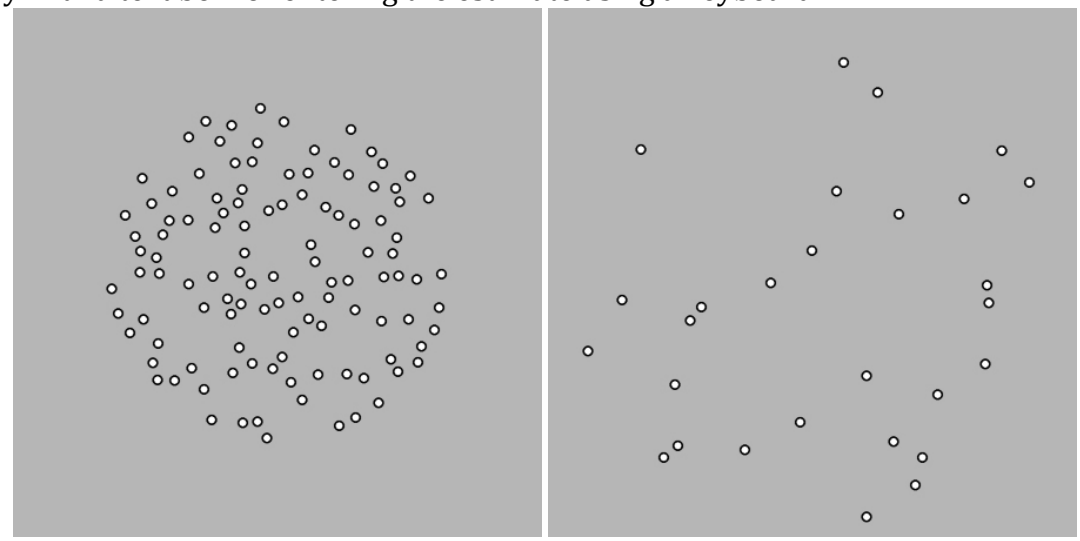


Figure 1. Sample numerosity displays using luminance-balanced dots. The left display depicts the smallest area used (with 112 dots), while the right display (with 28 dots) depicts the largest area. Note that luminance balancing will not be perfectly preserved in these reproductions.

Design. The range of numbers presented was manipulated between subjects. For the full-range condition, a quasi logarithmic range of 32 different numbers starting from 1 was

created by including the first 10 integers and then increasing by 2 up to 20, by 4 up to 40, by 8 up to 80, by 16 up to 160 and by 32 up to 224: (1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 14, 16, 18, 20, 24, 28, 32, 36, 40, 48, 56, 64, 72, 80, 96, 112, 128, 160, 192, 224). The highest density used (224 dots presented in the smallest area) was 1.2 dots/deg^2. For participants in the upper-range condition, the top half of this range was densely sampled so as to also have 32 distinct values by inserting 16 intermediate numbers starting with 26: (26, 28, 32, 34, 36, 38, 40, 44, 48, 52, 56, 60, 64, 68, 72, 74, 80, 88, 96, 104, 112, 120, 128, 144, 160, 176, 192, 208, 224).

A block of 96 trials consisted of presenting the 32 distinct numerosities within each of the 3 sizes of circle. Two blocks of 96 randomly ordered trials were completed by each participant.

Procedure. After being informed of the general procedure, participants were given specific instructions to make estimates of the number of elements presented based on what they saw. They were encouraged to be as precise as possible in making their estimates and to avoid reporting only round numbers. They were then presented with the 192 experimental test trials one by one, typing in their estimate for each one. The experiment typically took about 20 minutes.

Analysis. To remove typographic errors, estimates that were equal to or less than the square root of the presented number (e.g., an estimate of 8 or less for 64 dots) were excluded and estimates that were more than twice the presented number were also excluded. These exclusions represented less than 1% of the data. Because numerosity estimation normally conforms to a power function, all statistics are reported for log-transformed data.

Results

Participants generally gave self-consistent responses: The median within-subject correlation between the log-transformed dot numbers and estimates was 0.94. A plot of the mean estimates is shown in Figure 2A; the mean within-subject coefficients of variation (CoVs), corrected for the small sample size (normally 6 estimates per subject for each number presented across all display sizes) are shown in Figure 2B. These graphs illustrate the existence of a number estimation elbow at about 20 dots and a corresponding elbow in the CoVs. The CoV data replicates the very low variance normally found in the subitizing range (up to about 4 dots in our data). This is followed by a rapid increase between 4 and 20 dots and a plateau starting at the location of the second estimation elbow.

A. B.

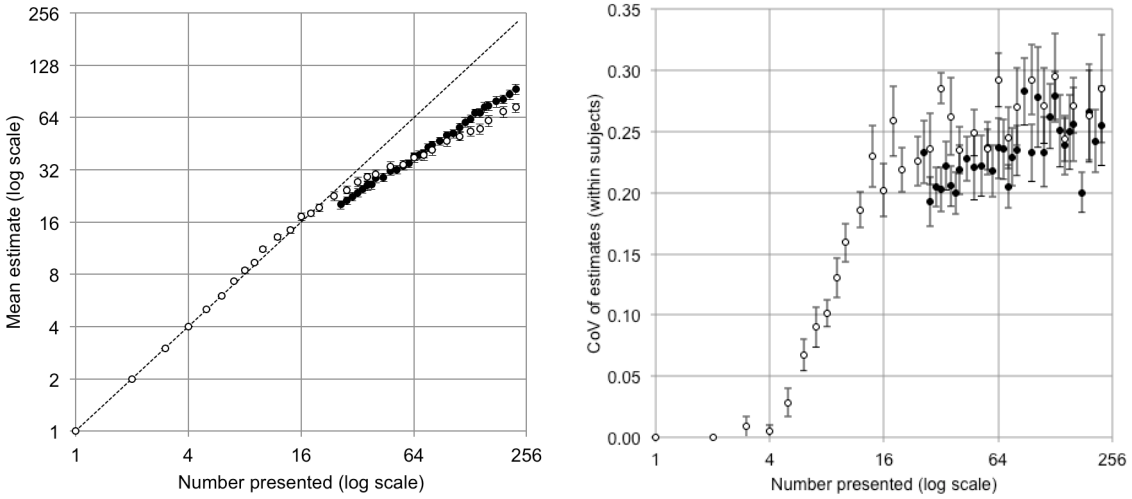
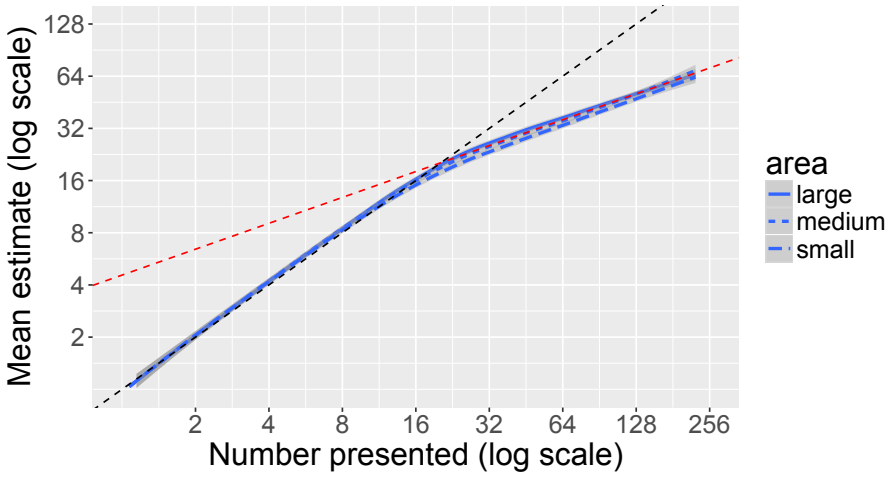


Figure 2. Plots of mean estimates (left) and mean within-subject CoVs (right) as a function of number presented separated by range of values estimates (filled circles for participants estimating in the upper range only), with between-subject standard error bars. For small numbers, errors bars are smaller than the plot points.

To provide a clearer picture of the estimation function, a plot of locally smoothed data (reflecting the non-independence of adjacent values) for the full-range condition is shown in log-log space in Figure 3A, as a function of display area, showing that the same elbow reported by Durgin (2016) is clearly present in the full-range data. A similar plot for the high-range only version of the experiment is shown in Figure 3B. The estimates from both groups in the upper range (28-224) are contrasted in Figure 4.

There are five principal features of the estimation data worthy of note: (1) There is an elbow in the full-range data (Figure 2A), showing a change in slope (i.e., exponent of the power function) that occurs somewhere between the numerosities of 16 and 32; (2) this elbow coincides with the start of an apparent plateau in the CoVs (Figure 2B), which rise dramatically between the subitizing range and the elbow; (3) Figures 3A and 3B suggest that both groups of participants replicate Krueger's (1972) observations of a power-function fit for numbers greater than 25 and both replicate his small effect of display area; however, (4) as shown in Figure 4, the exponents (log-log slopes) in the upper ranges differ across the two versions, consistent with greater sensitivity when only the upper range is tested. Moreover, (5) as suggested in Figure 2B, within-subject variability in the upper range (i.e., > 25 dots) is lower when only the upper range is estimated (Mean CoV = 0.23), than when the full range is tested (Mean CoV = 0.26), $t(46) = 4.31$, $p < .001$.

A.



B.

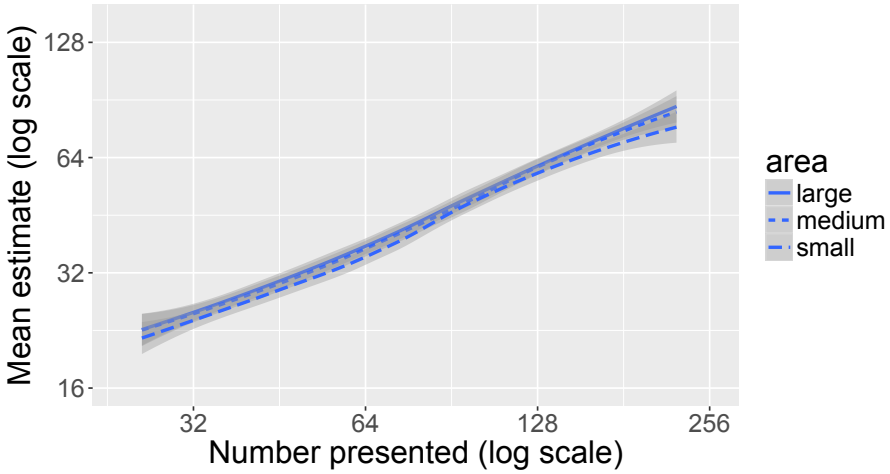


Figure 3. Log-log plots of estimates as a function of display area showing locally-smoothed estimates and between-subject standard errors of fit (medium gray). Panel A shows the estimation data from the full-range condition of Experiment 1, split by display area; superimposed on the data are a dotted black line representing accurate responding (exponent = 1.0), and a power-function fit (exponent = 0.50) to the upper portion of the data, illustrating the transitional elbow at a numerosity of about 20. Panel B show the estimation data from those who only estimated numerosities greater than 25.

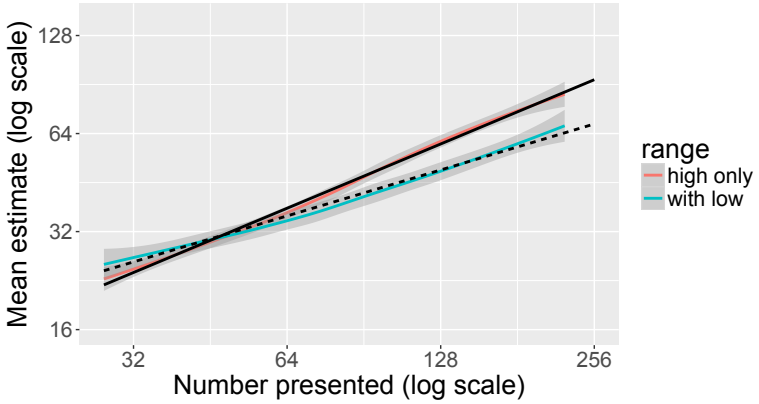


Figure 4. The effect of tested range on estimates in upper range in Experiment 1. Locally-smoothed estimates with between-subject standard errors for the upper range are shown as a function of range presented. The data from those who only saw numerosities greater than 25 (solid black fit line) were more steeply sloped than the data given by those who made estimates across a lower range as well (dashed black fit line).

To quantify the slope difference, the exponents of the best-fitting power-functions in the common range (greater than 25 dots) were calculated for each display area and participant (i.e., the slopes of best fits to the log-transformed values). An ANOVA on these exponents with display area (small, medium or large) as a within-subject factor and range (high or full) as a between-subject factor confirmed that the mean exponent in the high range group (M = 0.70, 95% CI [0.62, 0.77]) was reliably higher than the mean slope over the same numeric range in the full-range group (M = 0.50, 95% CI [0.45, 0.54]), $F(2, 76) = 0.27$, $p = .761$, $\eta^2_G = 0.32$. Note that an exponent of 0.5 is consistent with square root scaling of density (i.e., using information about linear separations among dots to estimate density), though other forms of explanation for this low exponent may be more likely (including poorer discrimination). The 0.5 exponent for the upper portion of the full range condition is somewhat lower than that observed by Huk and Durgin (1996; Durgin, 2016). The 0.7 exponent for the high range group, on the other hand, is consistent with that reported by Krueger (1972) for a similar range.

As the graphs in Figures 3A and 3B suggest, the ANOVA found no effect of display size on the exponents across observers, $F(2, 76) = 0.27$, $p = .761$. For the upper-range group, an ANOVA on log estimates indicated a highly-reliable effect of display size, $F(2, 38) = 11.5$, $p < .001$, $\eta^2_G = 0.06$, though the average difference from small to large was only 6.8% (*95% CI* [3.2%, 10.6%]) for a change in area by more than a factor of 2. For the upper range of values in the full-range condition (i.e., for numerosities 28 to 224), there was also a small, but reliable, effect of display size (M = 9.6%, 95% CI [4.4%, 15.0%]), $F(2, 38) = 12.5$, $p < .001$, $\eta^2_G = 0.15$.

Discussion

Our principal concern was whether, using central presentation, we would still see evidence of the elbow in the numerosity function at about 20 dots. Indeed, when numerosity estimates were made across a range including numbers from 1 to 224, there was a clear inflection in the number estimation function at about 20. In contrast, when the range was limited to values greater than 25, the produced estimation function is similar in detail to those reported by Krueger (1972), and still has a log-log slope much less than 1, similar to the estimation data beyond the elbow in the full-range data. Note that, as in

Kaufman et al.'s (1949) data, the exponent of the power function is about 1 in the low range (for 5-16) but is less than 1 in the high range. Thus even when large elements are used, display area is varied, and presentation is central rather than peripheral, we replicate the elbow reported by Durgin (2016; Huk & Durgin, 1996), and observable in the data of Kaufman et al.

This pattern of data suggests that different types of information may be used to estimate numbers above and below about 20. Although Figure 3A could be used to argue that the transition happens slightly earlier for smaller (more dense) display areas, Figure 3B, where no transition occurs, still shows an effect of display area. This suggests, instead, that the process involved in evaluating higher numbers is simply susceptible to a display-size bias, and this can account for the why there appears to be an earlier separation of the estimates in the smaller display areas. Although the effects of display size are superficially consistent with effects predicted by occupancy models, for example (reduced estimates of number for denser, smaller displays), the overall pattern observed here is not. Simulations of occupancy models show that the area bias should have a narrow window if it were based on a measure of filled area or occupancy (Allik & Tuulmets, 1991; Durgin, 1995), and would not produce an elbow of this sort.

Still, several questions remain. First, why is the slope for the upper range shallower for the participants making estimates for the lower range as well? It is possible that when only high numbers are shown, there is an advantage in discrimination for high numbers. If, in addition to subitizing processes, there are two different types of visual process involved in number estimation (i.e., above and below 20) and cognitive interfaces with these processes compete for resources, perhaps discrimination suffers for the upper range when working memory must keep cognitive interfaces for both kinds of process ready for application. This hypothesis is supported by the higher CoVs in the upper range among participants who judged both ranges.

Second, is the formation of the elbow dependent on the presence of the subitizing range? That is, might the fairly accurate estimation performance up to about 20 dots be parasitic on the presence of instances of numbers from the subitizing range to help calibrate estimation up to about 20? For example, being asked to exactly estimate five dots might enable one to better estimate ten dots, later, by calibrating a perceptual unit for "five" that can be scaled up when confronted with collections well above five dots. If this were true, then magnitude estimation for numbers below 20 might be less accurate if no small sets were shown during the experiment.

Experiment 2: Range independence

In Experiment 2 we sought to test whether the location of the visual number elbow depends on the inclusion of the subitizing range in the tested values. That is, it seemed possible that the subitizing range might serve to calibrate the estimation range (up to a point), such that the low numbers (i.e., below 7) help to scale the estimation of somewhat larger numbers (up to 20 or so). We therefore manipulated the range of numbers being estimated to test whether the presence or absence of the subitizing range affected the location or shape of the elbow. If the elbow seen in Experiment 1 merely represents a transition between subitizing-supported (scaled) estimation and self-scaled estimating, then eliminating the subitizing range might eliminate the elbow. Conversely, if the elbow

represents a division between two different kinds of estimation process, the transition should occur whether or not the subitizing range is included.

Methods

The stimuli, apparatus and general procedures were as in Experiment 1.

Participants. Forty-eight undergraduate students participated in partial fulfillment of a research requirement for the Introductory Psychology course. None had participated in Experiment 1. Half of the participants made estimates in each of the two ranges.

Design. The manipulation of numeric range (low – including the subitizing range vs. high – not including the subitizing range) was implemented between-subjects, with random assignment. Both groups of participants gave estimates of the elbow portion of the range used in Experiment 1. That is both groups saw 16 numbers that included the range from 9 to 72 (9, 10, 12, 14, 16, 18, 20, 24, 28, 32, 36, 40, 48, 56, 64, 72). The low-range (subitizing) group additionally gave estimates of all of the numbers from 1-8, intermixed with the shared range. The high-range (non-subitizing) group additionally gave estimates of 8 numbers above the shared range (viz., 80, 96, 112, 128, 144, 160, 192, 224). Thus each group had 8 numbers unique to that group in addition to the range of 16 numbers (half greater than 25 and half lower) spanning 9 to 72 that was common to both groups. As in Experiment 1, there were 3 different display sizes used, but only 24 distinct numbers, and thus a total of 72 trials constituted the full design. Two blocks of 72 randomly ordered estimation trials were collected.

Results and Discussion

If the presence or absence of the subitizing range was irrelevant to the formation of an elbow at about 20 elements, we should expect to see no effect of our manipulation of this range. The mean estimates for each condition are shown in Figure 5 against the fit lines computed for the full range data from Experiment 1. The figure suggests that when the subitizing range was excluded, there was an elevation of estimates in the upper range. Computing the log-log slope for each participant in this upper range (values greater than 25) produced a mean value of 0.61 (95% CI [0.54, 0.68]), which is higher than the slopes in this upper range for the full range condition of Experiment 1, $t(42) = 2.71$, $p = .010$, $d = 0.84$. This elevation of the exponents could have been due to the elimination of the subitizing range. But it might also be attributable to the greater proportion of trials that were in the upper range as a consequence of eliminating the subitizing range.

In order to test whether the effect of contextual range (high or low) differentially affected the upper range of numbers, an ANOVA was conducted on estimates over the shared range (9-72), with eight levels of number and two levels of magnitude (representing 9-24 vs. 28-72), as well as size as within-subject variables, and contextual range as a between-subject variable. Because there was a marginal interaction between number and magnitude (consistent with the presence of a discontinuity), $F(7, 322) = 392.4$, $p < .001$, $\eta^2 = 0.66$, separate ANOVAs were conducted for each magnitude level.

One ANOVA on the mean log estimates at each numeric value above 25 in the shared range of the present experiment used number (i.e., 28 to 72) and display area (3 levels) as within-subject factors, and contextual range (high or low) as a between-subject factor. It is of no surprise that log estimates varied as a function of number, $F(7, 322) = 112.6$, $p < .001$, $\eta^2 = 0.28$, and that the effect of display size was also reliable, $F(2, 92) = 20.0$, $p < .001$, $\eta^2 = 0.02$. What is most pertinent is that log estimates were higher for this range when the

contextual range included higher numbers rather than lower numbers, $F(1, 46) = 4.94$, $p = .031$, $\eta^2_G = 0.05$.

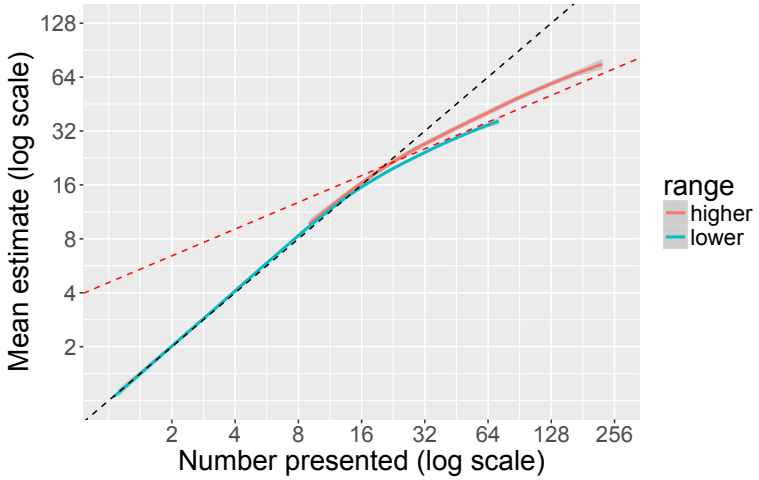


Figure 5. Results of Experiment 2. Mean estimates in log-log space are shown as a locally smoothed plot with between-subject standard errors (medium gray). The dashed lines represents actual number (black), as well as the fit line (red; log-log slope = 0.5) found for the upper range of the full range group in Experiment 1.

A second ANOVA on mean log estimates at each numeric value below 25 in the shared range of the present experiment used number (i.e., 9 to 24) and display area (3 levels) as within-subject factors, and contextual range (high or low) as a between-subject factor. Again, estimates varied as a function of number, $F(7, 322) = 392.4$, $p < .001$, $\eta^2 = 0.66$, and the effect of display size was also reliable in this range, $F(2, 92) = 15.7$, $p < .001$, $\eta^2 = 0.02$. However the estimates for these numerosities below 25 did not differ reliably as a function of the contextual range, $F(2, 28) = 2.28$, $p = .138$.

Discussion

Eliminating the subitizing range did not remove the discontinuity in the number estimation function that seems to occur at about 20 items. Despite the absence of a subitizing range, there still appeared to be two different types of estimation at play, and the transition between them occurred at about the same value. Eliminating the subitizing range did allow the slope of the log-log estimates above 25 to increase relative to the full-range condition of Experiment 1.

Experiment 3: Truncating the upper range

As a test of the stability of this discontinuity point, we sought to truncate the upper range of values to see if eliminating numbers above 40, for example, could eliminate the discontinuity. Perhaps numbers between 20 and 40 could be absorbed into the same process used to evaluate 10-20 if no higher numbers had to be evaluated during the same experimental procedures. To achieve a cap of 40 while retaining 24 distinct values, we sampled more densely from the range of 26-40. The experiment otherwise proceeded as before. Could restricting visual analysis to this range eliminate the shift toward the lower slopes associated with the upper range of numbers in the previous experiments? Alternatively, if the location of the elbow is impervious to the range of numbers used, it

may reflect a transition between two distinct visuo-cognitive processes for number estimation.

Methods

The methods were largely the same as in low-range condition of Experiment 2, and were designed to allow comparison to the lower range employed there (1-72). That is, the same number of estimates were collected and the analysis proceeded as before.

Participants. Twenty-four undergraduate students participated in partial fulfillment of a research requirement for the Introductory Psychology course. None had participated in Experiments 1 or 2.

Numeric range. Numbers from 2 to 40 were tested. In addition to 16 values between 2 and 24 (i.e., the eight numbers from 2 to 9, and the eight even numbers from 10-24), the upper portion of the range was now represented by the eight even numbers from 26 to 40.

Results

A plot of the mean log estimates is shown in Figure 6A against the fit line from Experiment 1. The same data are shown as a function of display size in Figure 6B. It is evident that capping the upper range at 40 did not substantially alter the estimates in this range. Mean estimates of the log-log slope for each participant for numerosities greater than 25 (M = 0.52, *95% CI* [0.38, 0.66]) did not differ from those in Experiment 2 in upper range (26-72) of the low-range condition (1-72), $t(47) = 0.12$, $p = .904$.

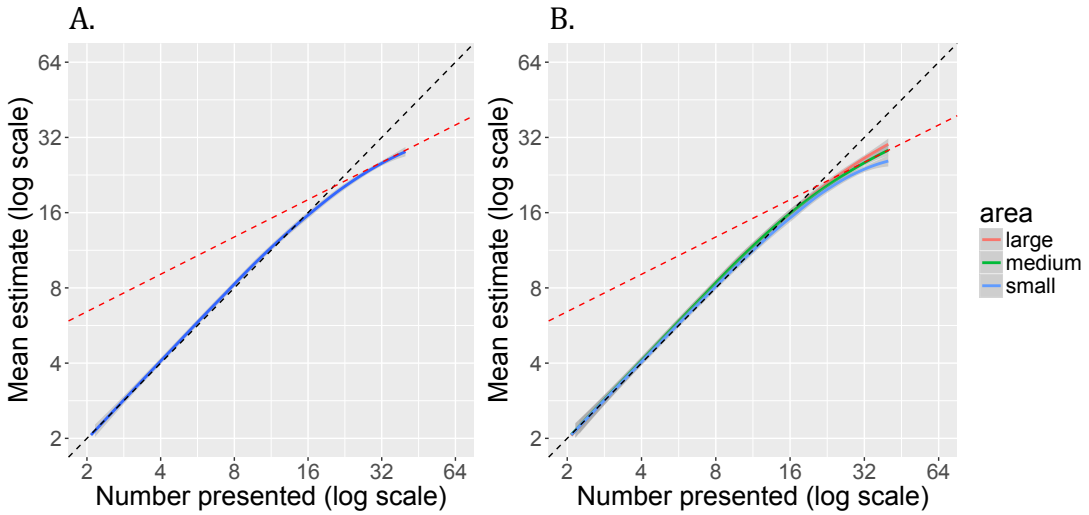


Figure 6. Results of Experiment 3. Mean estimates in log-log space are shown as a locally smoothed plot with between-subject standard errors (medium gray). A. The left panel shows the overall curve, collapsing across display size. B. The right panel shows the curves as a function of display size. The dashed lines in each graph represent actual number (black), as well as the fit line (red; log-log slope = 0.5) found for the upper range of the full range group in Experiment 1.

As in Experiment 1, there was a small size effect in the upper range. That is, a within-subject ANOVA on mean log estimates for numerosities greater than 25 (i.e., 26-40), demonstrated that estimates in the upper range differed by display area, $F(2, 48) = 11.5$, $p < .001$, $\eta^2_G = 0.02$. The magnitude of this effect was equivalent to an 11% increase in estimates from the smallest to the largest displays, which is similar to that found in Experiment 1 (i.e. 10%) for the upper range of the full-range condition.

Discussion

Experiment 3 showed that the location of the elbow seems stable even when the range of number values tested is shifted dramatically. Even with the subitizing range in place, the elbow remained in essentially the same location when the upper range was capped at 40, as when it extended to 224 (Experiment 1) or to 72 (Experiment 2). The evidence for this is that the very same overall fit lines drawn to capture the estimation data in Experiment 1 provided a good fit to the data from Experiment 3. Of course, as before, these two fit lines intersect at 20.

The display area effects (amounting to a 10% increase in estimates with a 125% increase in area) appear to emerge most clearly at the transition between mid-range numerosities (8-16) and the upper range (25 and up). This provides additional support for the idea that numerosity estimation is based on different visuo-cognitive processes in these two ranges. The effects of area may be consistent with an imperfect scaling of the integration of density and area information in the upper range.

General Discussion

Our goal in this paper has been to explore the effects of numeric range on the estimation of numerosity in order to establish the existence of a transition between two different types of estimation at about 20 dots. Across three experiments we tested ranges that started in the subitizing range (i.e., at 1 or 2), above the subitizing range (at 9) or in a range starting above 25. In these studies, we observed that the log-log slope of the numerosity estimation function above 25 was lowest (about 0.5) in the three experimental tests where the subitizing range was included, and highest (about 0.7) when all numbers tested were above 25. When estimation began at 9, the slope in the upper range was intermediate between these values (about 0.6). Whatever, the starting point, however, the location of the change in slope from about 1.0 to a much lower value was consistently at about 20 dots. Whether or not these particular slope values would generalize to other types of dots, the striking elbow repeatedly observed in our data replicates a discontinuity present in the data of Huk and Durgin (1996; Durgin, 2016), as well a Kaufman et al. (1949) using very different kinds of dots.

Much like the discontinuity at 6 that Kaufman et al. (1949) argued represented a break between what they labeled the subitizing range and the estimation range, the present discontinuity is both dramatic and stable across many studies. Although its precise cause is not understood, there is clearly a discontinuity in the way estimates are formulated beyond 20 dots. Human language number systems are usually base 10 or base 20, and this has been linked to the number of "digits" (fingers or fingers and toes) typically allotted to a human. The consistent emergence of the elbow at about 20 may therefore be related to the verbal number system itself somehow: From 5 up to 20, visual number can be estimated only approximately (i.e., with some noise) but there is relatively little bias in the estimates; there is a near one-to-one correspondence between presented numbers and the integers represented by the mean estimates. Beyond this range, estimation becomes increasingly compressed and one-to-one correspondence is therefore not maintained.

In this spirit, we propose to call the upper range the "superdigital" range of numeric estimation. By suggesting something beyond "digital" (i.e., beyond the typical number of fingers and toes) this term has the advantage of capturing both the transition point of the superdigital range (about 20), and the lowered exponent in the superdigital range: Because

the exponents in this range are consistently less than one, this range clearly cannot involve one-to-one correspondence (even approximately). That is, numerosities in this range are treated more like continuous magnitudes than discrete multitudes.

It seems unlikely that the discontinuity is just a matter of the numeric estimation process itself rather than an interaction between the estimation process and the perceptual processes that support it. This is because this discontinuity is not a typical feature of magnitude estimation using numbers. It seems reasonable to suggest that it may represent a transition to an approximate number domain where the most efficient means of estimating number involves combining density and area information (Dakin et al., 2011; Durgin, 1995, 2001).

Notably, however, the location of the discontinuity does not appear to be determined by density, per se, and thus may be unrelated to the density-based transition in number-comparisons proposed by Anobile et al. (2014). That is, the location of the elbow observed here clearly did not change location by a factor of 2 when the size of the display (and thus the density) was altered by a factor of 2.25, as would have to the case if the transition point were determined by display density. Moreover, Anobile et al. (2015) suggest that the transition in their experiment occurred at 2.3 dots/deg^2 in central vision, and 0.5 dots/deg^2 15° in the periphery. Anobile et al. (2014) suggest a transition at 0.25 dots/deg^2. However, the transition point of 20 dots in the present experiment would represent a density of only 0.11 dots/deg^2 in the smallest display and only 0.05 dots/deg^2 in the largest. Thus, the location of the second number elbow seems likely to have a different origin. After all, there is a fundamental difference between the estimation task studied here and the magnitude comparison tasks used by Anobile et al., in that only estimation tasks require establishing an integer value for a collection.

Nor does this transition point fit with the speculations of Laski and Siegler (2007) regarding developmental patterns. Rather, the lowered exponent in the superdigital range is consistent with a great deal of psychophysical work on the scaling of number itself (e.g., Banks & Hill, 1974). This work suggests that logarithmic encoding predominates in our conceptual representations of number, and that this is most evident for unbounded ranges (as in estimation). Bounded ranges (such as number lines) tend to produce more linear coding (Banks & Coleman, 1981). If the range up to 20 were considered bounded (because of an estimation process that can only manage up to about 20 elements), and the range beyond twenty unbounded, the elbow at 20 might represent that transition.

Perhaps the simplest theory of why approximately linear estimation breaks down at about 20 can be derived from applying a subitizing process iteratively. If 2 to 4 clusters of 3 to 5 dots can serve as a basis for estimating number with little bias in the range of 6-20, for example, the elbow we have observed here may be bounded by the upper limits of adding (or multiplying) together a subitizable number of roughly subitizable clusters. If some sort of grouping and adding strategy does distinguish the middle range from the superdigital range, we suspect such a strategy may depend on the presence of a linguistic system that represents numbers (e.g., Gordon, 2004) as well as fluency in basic arithmetic knowledge. For now we leave these speculations for future tests.

References

Allik, J., & Tuulmets, T. (1991). Occupancy model of perceived numerosity. *Perception, & Psychophysics, 49*, 303-314.

Anobile, G., Cicchini, G. M., & Burr, D. C. (2014). Separate mechanisms for perception of numerosity and density. *Psychological Science*, *25*, 265-270.

Anobile, G., Turi, M., Cicchini, G. M., & Burr, D. C. (2015). Mechanisms for perception of numerosity or texture-density are governed by crowding-like effects. *Journal of Vision, 15*(5), 4. https://doi.org/10.1167/15.5.4

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

Banks, W. P., & Coleman, M. J. (1981). Two subjective scales of number. *Perception & Psychophysics*, *29*, 95-105.

Banks, W. P., & Hill, D. K. (1974). The apparent magnitude of number scaled by random production. *Journal of Experimental Psychology*, *102*, 353-376.

Brainard, D. H., & Vision, S. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433-436.

Burr, D., & Ross, J. (2008). A visual sense of number. *Current Biology, 18*, 425-428.

Cohen, D.J. & Quinlan, P.T. (2018). The log–linear response function of the bounded number-line task is unrelated to the psychological representation of quantity. Psychonomic Bulletin & Review, 25: 447-454. https://doi.org/10.3758/s13423-017-1290-z

Carlson, C. R., Moeller, J. R., & Anderson, C. H. (1984). Visual illusions without low spatial frequencies. *Vision Research*, *24*, 1407 – 1413.

Dakin, S. C., Tibber, M. S., Greenwood, J. A., & Morgan, M. J. (2011). A common visual metric for approximate number and density. *Proceedings of the National Academy of Sciences*, *108*(49), 19552-19557.

Durgin, F. H. (1995). Texture density adaptation and the perceived numerosity and distribution of texture. *Journal of Experimental Psychology: Human Perception and Performance, 21*, 149-169.

Durgin, F. H. (2001). Texture contrast aftereffects are monocular; texture density aftereffects are binocular. *Vision Research, 41*, 2619-2630.

Durgin, F. H. (2016). Magnitudes in the coding of visual multitudes: Evidence from adaptation. In A. Henik (ed.) *Continuous Issues in Numerical Cognition: How Many or How Much?* (pp. 245- 270). Elsevier.

Durgin, F. H. & Huk, A. C. (1997). Texture density aftereffects in the perception of artificial and natural textures. *Vision Research, 37*, 3273-3282.

Durgin, F. H., & Li, Z. (2011). The perception of 2D orientation is categorically biased. *Journal of Vision, 11(8):13*, 1-10.

Fernberger, S. W. (1921). A preliminary study of the range of visual apprehension. *The American Journal of Psychology*, *32*, 121-133.

Gebuis, T., & Reynvoet, B. (2012). The interplay between nonsymbolic number and its continuous visual properties. *Journal of Experimental Psychology: General*, *141*, 642-648.

Gordon, P. (2004). Numerical cognition without words: evidence from Amazonia. *Science*, *306*, 496–499.

Hamilton, W. (1859). *Lectures on metaphysics and logic* (Vol. 1). Gould and Lincoln.

Huk, A. C., & Durgin, F. H. (1996). Concordance of numerosity comparison and numerosity estimation: Evidence from adaptation (abstract). *Investigative Ophthalmology & Visual Science*, *37*, S292.

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. *The American Journal of Psychology*, *62*, 498-525.

Kleiner, M., Brainard, D., & Pelli, D., (2007). What's new in Psychtoolbox-3? *Perception, 36* ECVP Abstract Supplement.

Krueger, L. E. (1972). Perceived numerosity. *Perception, & Psychophysics, 11*, 5-9.

Krueger, L. E. (1982). Single judgments of numerosity. *Perception, & Psychophysics, 31*, 175-182.

Laski, E. V., & Siegler, R. S. (2007). Is 27 a big number? Correlational and causal connections among numerical categorization, number line estimation, and numerical magnitude comparison. *Child Development*, *78*, 1723-1743.

Leibovich, T., Katzin, N., Harel, M., & Henik, A. (2017). From "sense of number" to "sense of magnitude": The role of continuous magnitudes in numerical cognition. *Behavioral and Brain Sciences, 40*, E164. doi:10.1017/S0140525X16000960

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

Taves, E. H. (1941). Two mechanisms for the perception of visual numerousness. *Archives of Psychology (Columbia University)*.

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.

Van Oeffelen, M. P., & Vos, P. G. (1982). A probabilistic model for the discrimination of visual number. *Perception & Psychophysics*, *32*, 163-170.