

Spatial number estimation has a higher linear range than temporal number estimation; differential affordances for subdivision might help to explain why

Frank H. Durgin

Department of Psychology, Swarthmore College,
Swarthmore, PA, USA



Elsie Aubry

Department of Psychology, Swarthmore College,
Swarthmore, PA, USA



Julius J. Balisanyuka-Smith

Department of Psychology, Swarthmore College,
Swarthmore, PA, USA



Çiçek Yavuz

Department of Philosophy, Haverford College,
Haverford, PA, USA



Estimation of visuospatial number typically has a limited linear range that goes well beyond the subitizing range but typically not beyond 20 items without calibration procedures. Three experiments involving a total of 104 undergraduate students, each tested once, sought to determine if the limit on the linear range represented a capacity limitation of a linear accumulator or might be the result of a strategy based on subdividing spatial displays into potentially subitizable subsets. For visual and auditory temporal numbers for a large range of numbers (2–58; Experiment 1), the (unbiased) linear range was found to be quite restricted (three or four items). Using matched linear spatial number stimuli (Experiment 2), the linear range observed extended to about nine or 10 items. Experiment 3 compared estimates when simultaneous two-dimensional spatial number displays were presented briefly, with estimates for identical displays that accumulated over time. The linear range of estimates for accumulating spatial displays reached only 11 items, whereas that for briefly presented displays extended to about 20 items. These results suggest that the limit on the linear range is not simply a capacity limitation in a linear accumulator. Rather, they support the idea that linear spatial number estimation for the range from five to 20 may be based on subdividing the display into a subitizable number of (potentially) subitizable groups, even if those groups are not outwardly marked.

Introduction

In the absence of pretraining or other opportunities to calibrate spatial number estimation, Portley and Durgin (2019) observed that estimates of visual number were, on average, accurate up to about 20 items but underestimated actual number thereafter. By varying the range of numbers tested, Portley and Durgin showed that this effect was not due to some sort of central tendency but instead seemed to occur near 20 even when relatively few numbers beyond 20 were tested and most were below, or when relatively few numbers below 20 were tested and most were above 20. In all cases, estimates in the range up to 20 were roughly linear and accurate, whereas the range beyond 20 was nonlinear (a power function) and systematically underestimated number (had an exponent less than 1). Power functions have long been observed for estimation of large visual numbers (Krueger, 1972; Krueger, 1984). In the present study, we examined temporal number estimation in order to test two alternative hypotheses about the apparent limit at 20.

One hypothesis, put forward by Portley and Durgin (2019), is that 20 represents a kind of natural limit of a spontaneous grouping process that might seek to divide the dots into four or five clusters—a strategy that might lead to accurate estimation if each cluster were only about four or five dots or fewer. Subitizing processes (Kaufman et al., 1949) are thought to be

Citation: Durgin, F. H., Aubry, E., Balisanyuka-Smith, J. J., & Yavuz, Ç. (2022). Spatial number estimation has a higher linear range than temporal number estimation; differential affordances for subdivision might help to explain why. *Journal of Vision*, 22(11):15, 1–11, <https://doi.org/10.1167/jov.22.11.15>.



limited to about four or five items. In other words, spatially random dot patterns tend to form clumps, and a subitizable number of natural clumps that are themselves roughly subitizable might be used to estimate visual numbers up to about 20 without substantial bias. On this view, unbiased estimation up to 20 is an extension of the subitizing range by means of active grouping, and 20 simply represents the point at which this strategy breaks down. Several recent studies have used explicit grouping information to enhance (reduce variance in) number estimation in the range below 20 (Ciccione & Dehaene, 2020; Starkey & McCandliss, 2014). Could implicit or automatic spatial grouping-by-subdividing processes underlie linear performance up to 20?

A version of this idea, which we will call *subdivision*, would be to suppose that number estimation in the range is often based on estimation of a subset (e.g., Solomon & Morgan, 2018) of possible clusters and might proceed by first dividing a whole array into a small number of roughly similar regions and then estimating some of those regions to arrive at an overall estimate. As long as the dots in most of those regions can be subitized, estimates might remain linear. If this kind of subdivision is an important part of extending linear estimation up to 20 for spatial number, we would expect that temporal number estimation might show a much smaller linear range. This is because it seems likely that ongoing temporal events whose terminal boundary is not known until the event is no longer present cannot be subdivided into a limited number (e.g., four or five) of roughly equal groups until it is too late to make effective use of the subdivision process (i.e., when the event is over).

An alternative hypothesis we considered was that 20 might represent the capacity of a linear accumulator such as the one discussed by Whalen, Gallistel, and Gelman (1999). Most sensory dimensions seem to be coded logarithmically. Logarithmic coding has the functional advantage of allowing for a very large range to be encoded with noise proportional to value and is thus an efficient coding scheme for many purposes. Moreover, logarithmic coding permits multiplication by means of addition. However, if a linear range is desired for the purpose of facilitating, say, addition or subtraction of linear quantities, then a coding scheme might need to be range limited in order to efficiently encode linear values. Whalen et al. (1999) developed a novel test of temporal number estimation in order to test whether a linear accumulator (with scalar variability) was a good fit for temporal number estimation. Specifically, whereas prior studies had used fixed rates to present temporal numbers, Whalen et al. used a stochastic process to produce sequences of temporal visual numbers ranging from 7 to 25. They reported that linear fits to their estimation data were excellent.

We had some prior reason to doubt that temporal number and spatial number were similar up to about 20. Whereas Whalen et al. (1999) observed a constant coefficient of variation (CoV) for number estimates for seven to 20 flashes, Portley and Durgin (2019) observed that within-subject CoVs for spatial number estimation increased gradually from four dots (the subitizing range), where they were nearly zero, up to about 16 dots, where they plateaued. Nonetheless, it remained to be seen whether and where temporal number estimates might show a transition between an unbiased linear range and a power function with an exponent less than 1.

In order to be able to tell whether there was a discontinuity in temporal number estimation at about 20, it seemed best to test a range of number that extended somewhat further beyond 20 than 25. For this reason, we adapted the method used by Whalen et al. (1999) but used a larger range of numbers. This, we reasoned, should reveal whether the discontinuity in spatial number at 20 observed by Portley and Durgin (2019) was also present for temporal number. We performed three replications of this basic experiment in **Experiment 1**: once with visual events, once with auditory events, and once with combined audiovisual events. In each case, we observed evidence that temporal number was not linear beyond a few items.

Experiment 1: Temporal number estimation for stochastic events

To test whether there was evidence of linear accumulation beyond about 20 temporal items, we asked participants to estimate the number of events presented in a stochastic manner like that used by Whalen et al. (1999). We tested visual events, auditory events, and combined audiovisual events among the participants. In most cases, event presentation was too rapid for subvocal counting (although, subjectively, counting of small numbers of auditory events was possible using echoic memory).

Methods

This and all subsequent experiments adhered to the tenets of the Declaration of Helsinki and were approved by the Swarthmore College institutional review board. All data collected for this study are available at the Open Science Foundation (https://osf.io/j7y3f/?view_only=8a20e97391334697903da4999f708dc2).

Participants

Forty-eight undergraduate students 18 years old or older at Swarthmore College participated as part

of their Introductory Psychology class. None had previously participated in a number estimation task. Sixteen participants estimated numbers of visual events, 16 estimated numbers of auditory events, and 16 were presented with audiovisual events for estimation.

Apparatus

The experiment was coded using PsychToolbox (Brainard, 1997; Kleiner, Brainard, Pelli, Ingling, & Broussard, 2007) running in MATLAB (MathWorks, Natick, MA) on a Mac Pro (Apple, Cupertino, CA) with a 120-Hz VIEWPixx/EEG display (VPixx Technologies, Saint-Bruno, QC, Canada).

Design and stimuli

Twenty numbers were selected for the test using a base multiplier of 2.49 and an exponential multiplier 1.18 raised to the integer powers of 0 to 19; the resulting numbers were rounded to the nearest integer. These integers ranged from 2 to 58. Each visual event consisted of a circular spot (40 pixels across, about 2° in diameter) on the otherwise white screen being turned black for three frames (25 ms) and then white again for six frames (50 ms). Each auditory event consisted of a 25-ms click followed by 50 ms of silence. Two different stochastic constants were used: On each successive frame (8.33 ms), the probability of another event commencing was $1/k$, where k was 10 for half the trials and 14 for half the trials. The experimental trials consisted of three blocks of 40 trials, with each block consisting of a randomly ordered presentation of the combination of each of the 20 numbers with each of the two stochastic values. Empirically, the mean duration for 58 events was 9.70 s ($SD = 1.14$), with a mean of 10.63 s ($SD = 0.74$) for $k = 14$ and a mean of 8.79 s ($SD = 0.59$) for $k = 10$. This suggests that the average rate of presentation averaged about 5.9 Hz overall (6.4 Hz and 5.4 Hz for the two constants, respectively). The maximum rate possible, given the algorithm, was 13.3 Hz.

Procedure and practice

On each trial, the onset of the temporal number stimulus was preceded by the white screen turning green for 500 ms and then returning to white for a randomly jittered period of 500 to 750 ms prior to the onset of the stochastic process that presented a predetermined number of events. After the final event in the sequence there was a 500-ms pause and then the screen turned yellow to signal that an estimate should be given. Participants typed their estimates into an editable textbox and pressed return when complete. Using exclusion rules developed for typographic errors (typos) by Portley and Durgin (2019), trials with estimates that

were less than the square root of the actual number presented were treated as typos and shuffled back into the remaining trials for the block to be repeated; trials with estimates greater than three times the actual number were also excluded from analysis.

To accustom participants to the procedure, there were seven practice trials using numbers that included three and 58. This was done to ensure that participants were exposed to the highest numbers to be tested, as well as to what we assumed would be an easy number. No feedback or explicit information about the range of numbers tested was given. After reading instructions about how the task would proceed, the participants were invited to begin the practice trials while the experimenter was in the room with them. After a few practice trials were complete and the participants seemed comfortable with the procedure, the experimenter left the testing room and closed the door. The experiment proceeded thereafter, stopping once to alert participants that the practice trials were over and again after two blocks of experimental trials to suggest that the participants take a brief break before completing the final third of the experiment. The entire procedure took participants about 20 minutes to complete.

Results

For each participant, the geometric mean estimate (mean of log estimates) and the CoV (standard deviation of log estimates) were computed for each presented number. The grand geometric means of the estimates for each event type are plotted in Figure 1 in log-log space. For visual flashes, t -tests conducted at each point in log space suggested that only for three events was there no evidence of systematic error in estimation. A power function with an exponent of 0.826 provided an excellent fit to the estimation data ($R^2 = 0.999$) for the full range of values. This result suggests that the coding of temporal number is nonlinear. Like Whalen et al. (1999), we observed approximately constant (scalar) variability for estimation of visual flash numbers, even down to as few as three visual events.

For auditory events, performance was unbiased for two and three events but reliably underestimated four or more events. A power function fit to the biased range had an exponent of 0.824 ($R^2 = 0.999$). Variability for auditory events was proportionally smaller for smaller numbers than it was for visual events, perhaps because of the possibility of using echoic memory (Craik, 1969; Neisser, 1967) to count small numbers of events.

For audiovisual events, t -tests provided no evidence of systematic error in estimation for two, three, or four events, and estimates for four audiovisual events were reliably higher than those for four auditory events,

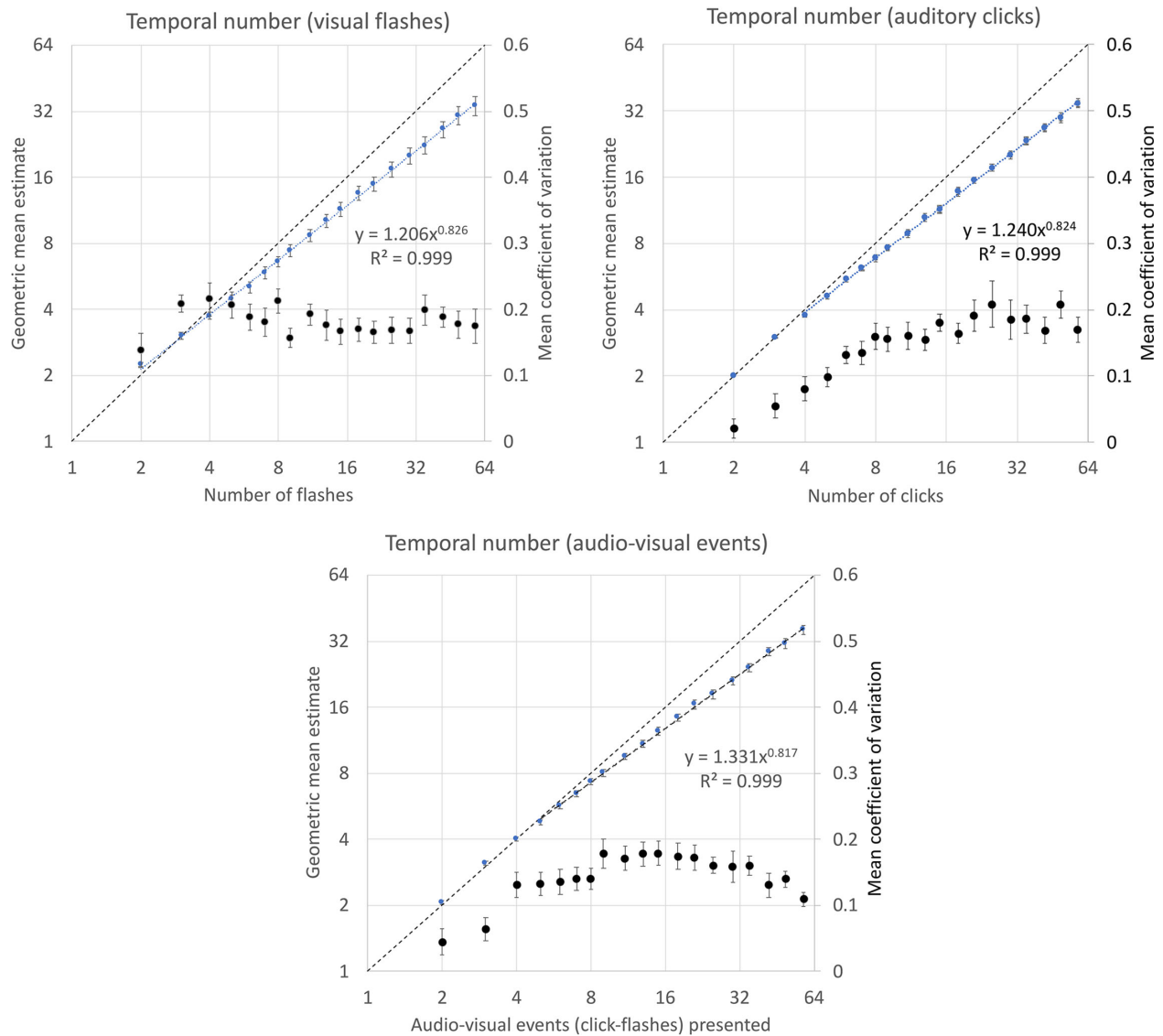


Figure 1. Temporal number estimation results of [Experiment 1](#) with flashes, clicks, or both are shown with power functions fit to the range demonstrating bias in each condition. Error bars represent standard errors of the means. Black disks represent mean CoVs (refer to right axis labels).

with $t(30) = 2.24$ and $p = 0.033$. However, there was systematic underestimation for five or more events. A power function fit to the range of five to 58 events had an exponent of 0.817 ($R^2 = 0.999$). Variability (CoVs) appeared to be lower for smaller numbers up to about eight.

Discussion

Our question was whether temporal number estimates would show an inflection at 20 as do spatial number estimates in the absence of training. Unlike estimates of spatial number, estimates of temporal number showed only a very limited linear (unbiased) range (see also [Philippi et al., 2008](#)). It thus seems

unlikely that the inflection at 20 for spatial number is the result of a capacity limitation in a linear accumulator.

Experiment 2: Number estimation for linearly arranged spatial number

Typical spatial number arrays afford two-dimensional (2D) subdivision by being spread over a 2D space. In contrast, temporal arrays are essentially one dimensional (1D). As a fairer comparison between spatial and temporal numbers, we therefore tested spatial number estimation for briefly presented 1D

spatial number arrays that were designed to mimic the temporal arrays of Experiment 1. Because these spatial displays afford subdivision/grouping only along one dimension, we anticipated that they might have a smaller range of unbiased performance than do 2D spatial arrays. However, even 1D spatial number arrays should better afford strategic subdivision simply because the entire array is presented at once. Thus, if a subdivision strategy is effective even for 1D arrays (as long as they are presented all at once, spatially), we should expect to see a larger linear range for linear spatial number in the present experiment than we did for temporal number in Experiment 1.

Method

Participants

Twenty-four undergraduate students 18 years old or older at Swarthmore College participated online (due to COVID-19 restrictions) as part of their Introductory Psychology class. None had previously participated in a number estimation task.

Displays and design

The experiment was administered using PsyToolkit (Stoet, 2010; Stoet, 2017). Number displays were generated in advance using Psychtoolbox (see Figure 2). The design was identical to that used in the temporal number experiments, except that horizontal position was substituted for time. Indeed, the visual images were generated from randomly selected trials from Experiment 1 by converting temporal intervals into spatial ones. A caveat is that, in order to render the most numerous stimuli within an 800-pixel-wide window, the spatial resolution of our



Figure 2. A sample of 1D spatial number stimuli from Experiment 2. Each row of vertical lines is also a spatial representation of a temporal number stimulus from Experiment 1. The numbers shown here, top to bottom, are 6, 9, 13, 21, 35, and 58 (twice), originally generated with alternating probability constants (1/14 for odd rows, 1/10 for even rows).

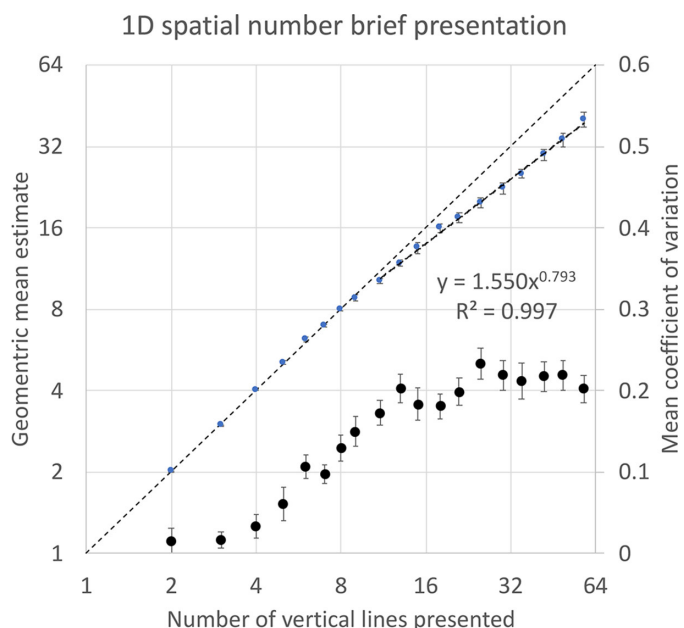


Figure 3. Spatial number estimation results for linear arrays of elements. Black disks represent the mean CoVs (scale on right side of graph). Error bars represent standard errors of the means. Fit line is applied only to estimates in the range showing statistical evidence of underestimation (11–58).

images was less than the temporal resolution had been, such that 4 pixels substituted for nine temporal frames. As a result, the vertical elements, at 2 pixels wide, were (proportionally) 50% wider than the durations of their temporal counterparts. Two versions of the experiment (each using a different set of 127 images) were created. Some sample images are shown in Figure 2. The display duration was 400 ms. Typos were excluded prior to analysis using the same criteria as in Experiment 1.

Results

A plot of estimates is shown in Figure 3. Performance with briefly presented linear spatial arrays was unbiased up to about nine items, which is clearly much better than for temporal arrays (Experiment 1). The power function fit to the range of underestimation has an exponent (0.79) that is quite similar to the exponents observed for temporal number stimuli in Experiment 1 (~0.82). The CoVs increased beyond the subitizing range, plateauing at about 11 or 13 elements.

Discussion

Because temporal events are unbounded until they end, the selection of a limited number of appropriate subsets for estimation cannot be easily undertaken until

after the last event has already occurred. Here we tried to use spatial displays that were otherwise analogous to the temporal number stimuli in [Experiment 1](#). We reasoned that even 1D (linear) visual number arrays likely afford efficient subdivision strategies that are difficult to deploy for temporal number. Although the range of unbiased estimation with these linear spatial arrays was more limited than that typically observed for 2D spatial number arrays, it far exceeded the unbiased range we saw for temporal number in [Experiment 1](#).

Though clearly speculative, it might be worth noting that studies of subitizing sometimes suggest that accuracy is limited to two or three linear items when they are not evenly spaced (e.g., [Krajcsi, Szabó, & Mórocz, 2013](#); [Mandler & Shebo, 1982](#)) or are not widely spaced ([Atkinson, Campbell, & Francis, 1976](#)). Thus, the reduced accuracy limit of nine items for the linear arrays used here is consistent with, for example, subdivision of our linear arrays into three groups, with accuracy attained for only up to three lines per group.

Experiment 3: Number estimation for spatially cumulative audiovisual events and for briefly presented spatial arrays

In [Experiment 1](#) we tested whether temporal number estimation seemed to reflect a linear accumulator and what the capacity limits of that accumulator were. The best estimate of unbiased performance (four events) occurred for audiovisual events. These results contrast with the observations of [Portley and Durgin \(2019\)](#) regarding estimation of spatial number, which is often unbiased up to about 20 items, even without calibration. In [Experiment 2](#), we used spatial displays where number was represented along a single (horizontal) dimension to better match one constraint of temporal number stimuli. We still found a larger range of unbiased performance than we observed for temporal number.

So far, then, it would appear that a subdivision or grouping hypothesis, like that proposed by [Portley and Durgin \(2019\)](#), seems more likely to explain unbiased performance up to about 20 items. Strategies that involve subdividing a display into a small number of subsets of elements can be efficiently exercised only when the entire array is present and are therefore easiest when all dots are presented simultaneously. In the present experiment, we sought to contrast a brief simultaneous presentation of spatial arrays with temporally extended presentations in which the same spatial number arrays, rather than appearing all at once, accumulated one dot at a time in a temporally

stochastic manner like that used in [Experiment 1](#). It would seem that the information content in a gradually accumulating display is no less than in the same display briefly presented, but if a subdivision strategy can only be implemented once the whole display is present then we should expect a strategic cost for a spatial display whose temporal bound, like that of a temporal number sequence, is unknown until it is too late.

Method

Thirty-two students who had not been in previous number estimation experiments participated. The temporal algorithm for the cumulative displays was identical to the one used in [Experiment 1](#). The difference was that each event was now composed of an auditory click, synchronized with the visual onset of a single element (black ring with an inner diameter of 12 pixels and an outer diameter of 18 pixels, about 1°) that remained on (and thus accumulated into a spatial array of elements) at random positions that were at least 24 pixels, center to center, from other elements within an unmarked circular boundary. The boundary was 600, 750, or 900 pixels in diameter to decouple density and number. The elements all disappeared 500 ms after the last one was added. Sixteen students participated in the cumulative version, and, for comparison with the simple spatial case, an additional 16 students made estimates of purely spatial patterns (presented for 400 ms), generated by the same spatial algorithm as that used in the cumulative case. Recall that, for the largest number presented (58), the mean accumulation time would have been about 10 s, but as with temporal number participants could not predict when the accumulation process would end. Moreover, during the accumulation process, the number of dots in any spatial subregion could increase at any time. Thus, although the final glimpse of the completed accumulated display lasted slightly longer than the brief, simultaneous display, participants probably could not know that it was the final display until it had disappeared.

Results and discussion

The grand geometric means for estimates in the two conditions are plotted in [Figure 4](#) in log–log space. For briefly presented spatial number arrays, we replicated the observation of [Portley and Durgin \(2019\)](#) that estimates were relatively unbiased up to about 20. In this case, there was some overestimation below 20, but the first evidence of underestimation occurred for 25 dots. In contrast, for the cumulative displays, unbiased performance was evident up to only 11 dots, with reliable underestimation first occurring for 13 dots. It is worth noting that estimates for the highest number

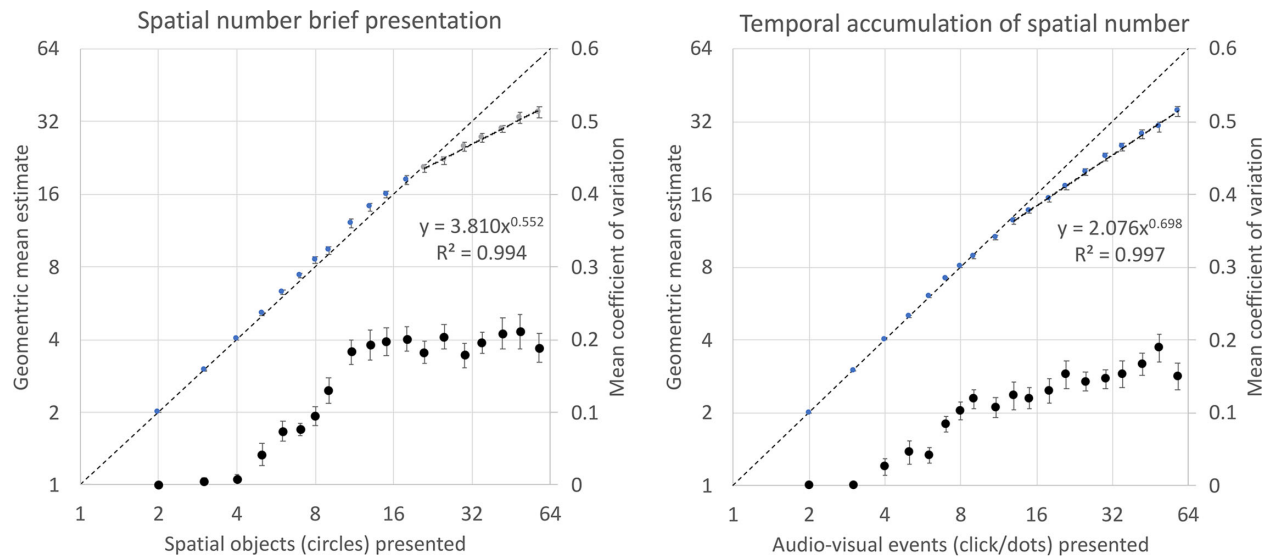


Figure 4. Number estimation results of [Experiment 3](#) with simple spatial arrays (left) or with multimodal events consisting of auditory clicks and visual onsets that accumulated into spatial arrays (right). Black disks represent the mean CoVs (scale on right side of graph). Error bars represent standard errors of the means. Fit lines are applied only to estimates in the range showing statistical evidence of underestimation (25–58 for brief spatial displays; 13–58 for temporally cumulative spatial displays).

were the same across both conditions (i.e., 58 dots were estimated as 35), which is similar to the estimates for 58 in [Experiment 1](#) (34, 35, and 36).

The rather large difference in linear range between the briefly flashed 2D spatial array and the gradually accumulating 2D spatial array is consistent with the idea that subdivision is used to produce accurate estimates up to about 20 for simultaneous spatial arrays. When spatial arrays emerge over an unpredictable time window, the spatial subdivision process cannot be as effectively wielded. Presumably this is because new dots were appearing up until the accumulation ended, whereas the end of accumulation was only signaled by the disappearance of the entire array (i.e., could not be easily anticipated).

Note that the within-subject CoVs show that, in the range above 10, the precision (CoV) of the estimates was better (smaller) for the gradually accumulated displays ($M = 0.14$) than for the brief simultaneous displays ($M = 0.19$), with $t(30) = 3.7$ and $p < 0.001$, despite the estimates being more accurate from 15 to 21 dots in the simultaneous condition. This seems consistent with the idea that the linear range reached 20 as a result of a cognitive strategy (e.g., involving subdivision) rather than more precise number perception, per se.

Although possible, in principle, we suspect that trying to repeatedly estimate the current state of the changing display by means of subdivision throughout the accumulation process would probably have been so labor intensive that participants would likely have depended on an alternative strategy to take advantage of both visual–spatial and auditory–temporal information.

The fact that the accumulation condition has a larger linear range than the audiovisual temporal condition of [Experiment 1](#) has many possible interpretations, but it could, for example, reflect a reduction in memory demand. Perhaps participants tended to use a hybrid temporal/spatial process that could combine low-variability estimates of up to about six or seven spatially accumulated visual elements, with temporal estimates of up to three or four new audiovisual events to arrive at unbiased estimates up to about 11.

The main conclusion we wish to draw from these results concerns the comparison of the cumulative condition to the brief, simultaneous spatial number conditions. The superior range of accuracy for brief, simultaneous presentation is consistent with the idea that, for a subdivision strategy to be accurate over the largest possible range, it should be initiated when all items are present.

General discussion

For numeric estimation of briefly presented spatial arrays of elements, estimates are largely unbiased up to about 20 ([Portley & Durgin, 2019](#)). Here, we considered two hypotheses about this observation and found greater support for the subdivision hypothesis than for the linear accumulator capacity limitation hypothesis. Specifically, for spatial numbers, the linear (unbiased) range of estimation was greater when all of the elements

were simultaneously but briefly presented than when the same spatial arrays accumulated over several seconds with accompanying auditory signals. The cumulative procedure produced a much larger unbiased range (up to about 11 elements) than temporal audiovisual signals with the same temporal statistics in [Experiment 1](#) (up to about four items), but a much lower range than 20 items for briefly but simultaneously presented spatial number arrays with the same spatial statistics. Although the unbiased range was reduced to nine items when linear simultaneous spatial arrays were used, this still far exceeded the performance for temporal number. These observations are consistent with the consideration that spontaneous subdivision strategies work best when all of the elements to be estimated appear at once so that an appropriate subset (e.g., one quarter of the elements) can be identified for initial analysis.

Prior temporal number data re-examined

Prior reports have suggested that temporal number estimation might be linearly related to presented number up to about 20 or 25 events ([White, 1963](#); [Whalen et al., 1999](#)). In contrast, we found here that temporal number estimates consistently followed a power function with an exponent of about 0.82 for numbers of events beyond about four or five. This observation led us to conduct a re-examination of some older data on temporal number estimation. By using data from published tables or, in some cases, digitizing

and extracting coordinates of data points from figures in these papers, we were able to reconstruct average estimates across five studies that did not use calibration procedures ([Cheatham & White, 1952](#); [Cheatham & White, 1954](#); [Forsyth & Chapanis, 1958](#); [Garner, 1951](#); [Lawrence, 1971](#)) and compared linear and power function fits to the data. Although all of the papers we examined reported linear fits, power functions generally provided equivalent fits for their data. [Table 1](#) shows the exponents for all five papers on temporal number estimation data as a function of modality and frequency of presentation.

[Figure 5](#) shows the exponents from [Table 1](#) as a function of presentation frequency with separately fit log functions to the auditory and visual data. [Figure 5](#) shows that, although exponents tended to be somewhat higher for auditory stimuli, estimation of both auditory and visual temporal stimuli seemed to be well captured by power functions with exponents less than 1 whenever the presentation frequency exceeded the rates for which subvocal counting was likely possible (i.e., 4 Hz). Moreover, the exponents appear to be systematically related to the presentation rate and modality.

Grouping theories and cognitive accumulation

The idea that grouping can contribute to number estimation has a long history. Although the term

Source	Modality	Range	Range fit	Frequency (Hz)	Exponent	R^2
Garner (1951)	Auditory	2–20	4–20	4	0.99	1.000
Garner (1951)	Auditory	2–20	4–20	6	0.96	1.000
Garner (1951)	Auditory	2–15	4–15	8	0.90	0.999
Garner (1951)	Auditory	2–12	4–12	10	0.87	0.997
Cheatham and White (1952)	Visual	1–9	4–9	10	0.80	0.986
Cheatham and White (1952)	Visual	1–13	4–13	15	0.75	0.980
Cheatham and White (1952)	Visual	1–18	4–18	22.5	0.62	0.942
Cheatham and White (1952)	Visual	1–20	4–20	30	0.67	0.957
Cheatham and White (1954)	Auditory	1–8	4–8	10	0.96	1.000
Cheatham and White (1954)	Auditory	1–11	4–11	15	0.76	0.996
Cheatham and White (1954)	Auditory	1–17	4–17	30	0.75	0.939
Forsyth and Chapanis (1958)	Visual	1–20	4–20	2.5	0.98	1.000
Forsyth and Chapanis (1958)	Visual	1–20	4–20	5	0.93	0.998
Forsyth and Chapanis (1958)	Visual	1–20	4–20	10	0.84	0.997
Forsyth and Chapanis (1958)	Visual	1–20	4–20	15	0.74	0.996
Forsyth and Chapanis (1958)	Visual	1–20	4–20	22.5	0.67	0.992
Forsyth and Chapanis (1958)	Visual	1–20	4–20	30	0.80	0.943
Lawrence (1971)	Visual words	1–15	4–15	4	0.99	0.993
Lawrence (1971)	Visual words	1–21	4–21	8	0.94	0.994
Lawrence (1971)	Visual words	1–31	4–31	12	0.86	0.994
Lawrence (1971)	Visual words	1–31	4–31	16	0.83	0.994

Table 1. Exponents computed for data from studies of temporal number estimation.

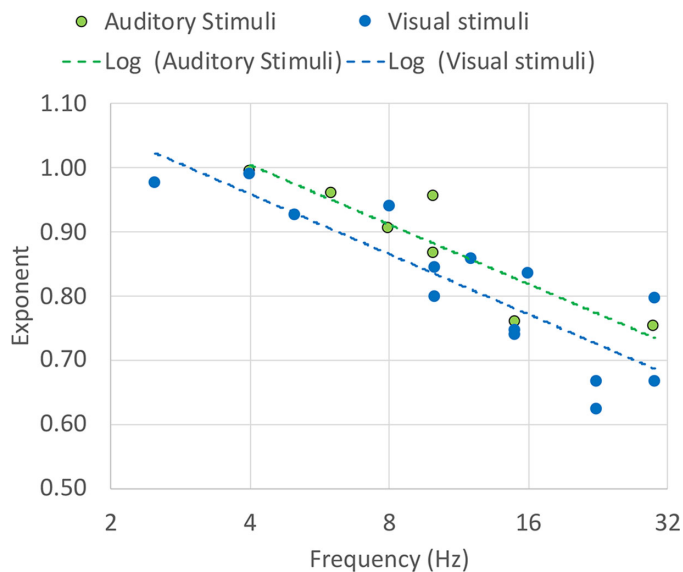


Figure 5. Exponents fit to temporal number estimation data from the five studies shown in Table 1 as a function of presentation frequency and modality.

groupitizing is relatively new (Starkey & McCandliss, 2014), Atkinson, Francis, and Campbell (1976) used Gestalt grouping principles for linear arrays (orientation, color, proximity) to test number estimation beyond the subitizing range (up to 12) and found that they could get errorless accuracy for up to eight items when they were split into two groups of four (i.e., two subitizable subsets). Somewhat later, Van Offelen and Vos (1982) reviewed the history of evidence for enumeration by subgroups and explored a model of visual processing that might select for small groups of items to be separately enumerated.

A large number of recent papers have supported the idea that Gestalt grouping (by color, proximity, etc.) supports improved number estimation—typically for numbers well below 20 (e.g., Maldonado Moscoso, Castaldi, Burr, Arrighi, & Anobile, 2020; Pan, Yang, Li, Zhang, & Cui, 2021). These grouping effects have even been found for auditory temporal number for up to about eight or nine items (Anobile, Castaldi, Maldonado Moscoso, Arrighi, & Burr, 2021). And evidence suggests that spontaneous group and add (or multiply) strategies can probably be deployed for simultaneous visual items defined in complex ways (by motion, for example; Kramer, Di Bono, & Zorzi, 2011).

Although it has been argued, from studies of dyscalculia, that groupitizing effects do not depend on mathematical ability, the empirical evidence presented by Anobile, Marazzi, Federici, Napoletti, Cecconi, and Arrighi (2022) may not be relevant to evaluating the possible role of math raised in our study for three reasons. First is calibration; the range of numbers they tested (5–10) was told to all the participants in advance,

so the study did not address the accuracy of estimation. Second is ambiguity of outcome; the measure of precision used in the work (CoV of estimates) showed that both groups gave more uniform estimates when clustered displays were presented. In general, although reduced variation in estimates can be interpreted as evidence of more precise numerical perception, it can also reflect better categorical recognition. Finally, in their study category recognition—in particular, the observed differences in the uniformity of estimates between clustered displays and random spreads—might reflect a difference between recognizable (clustered) patterns and confusable (random) stimuli (Krajcsi et al., 2013; Wolters, Van Kempen, & Wijnhuizen, 1987). For example, although randomized each time, the clustered number 7 used by Anobile et al. (2022) always consisted of two clusters of three dots and one additional cluster of 1 [3,3,1,0], whereas the number 8 could appear as one of three different breakdowns: [2,2,2,2], [4,4,0,0], or [3,3,2,0]. Consistent with the categorical recognition view, the improvement shown for seven dots when groupitized appears much greater than that for eight dots among both control participants and those with dyscalculia (Anobile et al., 2022) (Figure 3). Because the authors stated that the cluster breakdowns they used were ones that showed the “most robust” (p. 8) effects in their prior studies, it is possible that their clustered displays were particularly categorizable based on features other than total number (Krajcsi et al., 2013). On this view, spatial groupitizing, using only a few repeated cluster breakdowns, might not be a generalizable example of a subdivision strategy. Conversely, it would be quite surprising, from our perspective, if, without feedback or other information about the range of values being tested, the linear range of estimation for fully random, briefly presented spatial displays were to extend as high as 20 among those with dyscalculia (see Ashkenazi, Mark-Zigdon, & Henik, 2013).

The present work has provided evidence that differential affordances for efficient subdividing and subset estimation might underlie differential success at the various forms of number estimation tasks we have considered. For example, we suppose that efficient subdividing is easier when all elements are presented simultaneously because this allows greater potential for the proper division of elements into a subitizable number of (potentially subitizable) subsets. Support for this idea emerged from the comparison of linear spatial number estimation (Experiment 2) and temporal number estimation (Experiment 1). A fairly direct test of the idea was implemented in Experiment 3, where mean accuracy for brief simultaneous presentations of spatial number arrays seemed to far outstrip accuracy for gradually accumulated spatial number arrays with the same spatial statistics. We have noted that the temporal

unfolding of sequential events (when there is a sufficiently large range of possible numbers) leaves them unbounded and thus difficult to subdivide until it is “too late.”

Although we believe that we successfully manipulated the affordance for subdivision in [Experiment 3](#), it remains unclear how subdivision leads to linear estimation. It is possible that linear (accurate) estimates result from combining unbiased estimates of all resulting groups, as suggested by [Portley and Durgin \(2019\)](#), or by estimation based on a subset of groups, as might be suggested by [Solomon and Morgan \(2018\)](#). A subdivision strategy could include a step that seeks to estimate the elements in what seems like the most representative regions and to extrapolate from there, but we have not tried to address these questions here.

Whereas [Atkinson et al. \(1976\)](#) supposed that increased capacity of accurate estimation from grouping depended on having separate channels within which subitizing could occur, an addendum to their theory might be a cognitive accumulation process. That is, the participants in their experiments were well aware of the simple facts of addition and multiplication and could literally do the math. Our current hypothesis is that arithmetic knowledge is probably what allows subdividing to extend the (uncalibrated) linear range of estimation beyond the subitizing range, but only up to about 20. It could be that estimation and comparison performance on relatively small numbers beyond the subitizing range correlates with math skills because it can be improved by those skills ([Starkey & McCandliss, 2014](#)).

Conclusions

In the absence of a calibration procedure, number estimation for spatial and temporal numbers tends to show log scaling beyond a limited range. In our view, it is the extent of the limited linear range that requires explanation. In this paper we have sought to compare the extent of that limited range across a variety of procedures that included relatively high numbers (up to 58). We did not see evidence for a fixed capacity linear accumulator. For temporal number involving auditory signals, the uncalibrated range appeared linear only for fairly small numbers where echoic memory may have supported counting. For spatial number, the range of approximate linearity was highest when some form of subdivision strategy was most likely to be effective: simultaneously presented 2D displays rather than ones that gradually accumulated over an unpredictable duration.

Keywords: number, temporal number, grouping

Acknowledgments

Commercial relationships: none.

Corresponding author: Frank H. Durgin.

Email: fdurgin1@swarthmore.edu.

Address: Department of Psychology, Swarthmore College, Swarthmore, PA, USA.

References

- Anobile, G., Castaldi, E., Maldonado Moscoso, P. A. M., Arrighi, R., & Burr, D. (2021). Groupitizing improves estimation of numerosity of auditory sequences. *Frontiers in Human Neuroscience*, *15*, 687321, <https://doi.org/10.3389/fnhum.2021.687321>.
- Anobile, G., Marazzi, M., Federici, S., Napoletti, A., Cecconi, L., & Arrighi, R. (2022). Unimpaired groupitizing in children and adolescents with dyscalculia. *Scientific Reports*, *12*(1), 5629, <https://doi.org/10.1038/s41598-022-09709-5>.
- Ashkenazi, S., Mark-Zigdon, N., & Henik, A. (2013). Do subitizing deficits in developmental dyscalculia involve pattern recognition weakness? *Developmental Science*, *16*(1), 35–46, <https://doi.org/10.1111/j.1467-7687.2012.01190.x>.
- Atkinson, J., Campbell, F. W., & Francis, M. R. (1976). The magic number 4 ± 0 : A new look at visual numerosity judgements. *Perception*, *5*(3), 327–334, <https://doi.org/10.1068/p050327>.
- Atkinson, J., Francis, M. R., & Campbell, F. W. (1976). The dependence of the visual numerosity limit on orientation, colour, and grouping in the stimulus. *Perception*, *5*(3), 335–342, <https://doi.org/10.1068/p050335>.
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436, <https://doi.org/10.1163/156856897X00357>.
- Cheatham, P. G., & White, C. T. (1952). Temporal numerosity: I. Perceived number as a function of flash number and rate. *Journal of Experimental Psychology*, *44*(6), 447–451, <https://doi.org/10.1037/h0061192>.
- Cheatham, P. G., & White, C. T. (1954). Temporal numerosity: III. Auditory perception of number. *Journal of Experimental Psychology*, *47*(6), 425–428, <https://doi.org/10.1037/h0054287>.
- Ciccione, L., & Dehaene, S. (2020). Grouping mechanisms in numerosity perception. *Open Mind*, *4*, 102–118, https://doi.org/10.1162/opmi_a_00037.
- Craik, F. I. (1969). Modality effects in short-term storage. *Journal of Verbal Learning*

- and Verbal Behavior, 8(5), 658–664, [https://doi.org/10.1016/S0022-5371\(69\)80119-2](https://doi.org/10.1016/S0022-5371(69)80119-2).
- Forsyth, D. M., & Chapanis, A. (1958). Counting repeated light flashes as a function of their number, their rate of presentation, and retinal location stimulated. *Journal of Experimental Psychology*, 56(5), 385–391, <https://doi.org/10.1037/h0047974>.
- Garner, W. R. (1951). The accuracy of counting repeated short tones. *Journal of Experimental Psychology*, 41(4), 310–316, <https://doi.org/10.1037/h0059567>.
- Kaufman, E. L., Lord, M. W., Reese, T. W., & Volkman, J. (1949). The discrimination of visual number. *The American Journal of Psychology*, 62, 498–525, <https://doi.org/10.2307/1418556>.
- Kleiner, M., Brainard, D., Pelli, D., Ingling, A., & Broussard, C. (2007). What's new in Psychtoolbox-3? *Perception*, 36(14), 1–16.
- Krajcsi, A., Szabó, E., & Mórocz, I. Á. (2013). Subitizing is sensitive to the arrangement of objects. *Experimental Psychology*, 60(4), 227, <https://doi.org/10.1027/1618-3169/a000191>.
- Kramer, P., Di Bono, M. G., & Zorzi, M. (2011). Numerosity estimation in visual stimuli in the absence of luminance-based cues. *PLoS One*, 6(2), e17378, <https://doi.org/10.1371/journal.pone.0017378>.
- Krueger, L. E. (1972). Perceived numerosity. *Perception & Psychophysics*, 11, 5–9, <https://doi.org/10.3758/BF03212674>.
- Krueger, L. E. (1984). Perceived numerosity: A comparison of magnitude production, magnitude estimation, and discrimination judgments. *Perception & Psychophysics*, 35(6), 536–542, <https://doi.org/10.3758/BF03205949>.
- Lawrence, D. H. (1971). Temporal numerosity estimates for word lists. *Perception & Psychophysics*, 10, 75–78, <https://doi.org/10.3758/BF03214318>.
- Maldonado Moscoso, P. A., Castaldi, E., Burr, D. C., Arrighi, R., & Anobile, G. (2020). Grouping strategies in number estimation extend the subitizing range. *Scientific Reports*, 10(1), 14979, <https://doi.org/10.1038/s41598-020-71871-5>.
- Mandler, G., & Shebo, B. J. (1982). Subitizing: an analysis of its component processes. *Journal of Experimental Psychology: General*, 111(1), 1–22, <https://doi.org/10.1037/0096-3445.111.1.1>.
- Neisser, U. (1967). *Cognitive psychology*. New York: Meredith.
- Pan, Y., Yang, H., Li, M., Zhang, J., & Cui, L. (2021). Grouping strategies in numerosity perception between intrinsic and extrinsic grouping cues. *Scientific Reports*, 11(1), 17605, <https://doi.org/10.1038/s41598-021-96944-x>.
- Philippi, T. G., van Erp, J. B., & Werkhoven, P. J. (2008). Multisensory temporal numerosity judgment. *Brain Research*, 1242, 116–125, <https://doi.org/10.1016/j.brainres.2008.05.056>.
- Portley, M., & Durgin, F. H. (2019). The second number-estimation elbow: Are visual numbers greater than 20 evaluated differently? *Attention, Perception, & Psychophysics*, 81(5), 1512–1521, <https://doi.org/10.3758/s13414-019-01804-6>.
- Solomon, J. A., & Morgan, M. J. (2018). Calculation efficiencies for mean numerosity. *Psychological Science*, 29(11), 1824–1831, <https://doi.org/10.1177/0956797618790545>.
- Starkey, G. S., & McCandliss, B. D. (2014). The emergence of “groupitizing” in children’s numerical cognition. *Journal of Experimental Child Psychology*, 126, 120–137, <https://doi.org/10.1016/j.jecp.2014.03.006>.
- Stoet, G. (2010). PsyToolkit: A software package for programming psychological experiments using Linux. *Behavior Research Methods*, 42(4), 1096–1104, <https://doi.org/10.3758/BRM.42.4.1096>.
- Stoet, G. (2017). PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology*, 44(1), 24–31, <https://doi.org/10.1177/0098628316677643>.
- Van Oeffelen, M. P., & Vos, P. G. (1982). Configurational effects on the enumeration of dots: Counting by groups. *Memory & Cognition*, 10(4), 396–404, <https://doi.org/10.3758/BF03202432>.
- Whalen, J., Gallistel, C. R., & Gelman, R. (1999). Nonverbal counting in humans: The psychophysics of number representation. *Psychological Science*, 10(2), 130–137, <https://doi.org/10.1111/1467-9280.00120>.
- White, C. T. (1963). Temporal numerosity and the psychological unit of duration. *Psychological Monographs: General and Applied*, 77(12), 1–37, <https://doi.org/10.1037/h0093860>.
- Wolters, G., Van Kempen, H., & Wijnhuizen, G. J. (1987). Quantification of small numbers of dots: Subitizing or pattern recognition? *The American Journal of Psychology*, 100(2), 225–237, <https://doi.org/10.2307/1422405>.