A primarily serial, foveal accumulator underlies approximate numerical estimation

Samuel J. Cheyette[a,1], and Steven T. Piantadosi[a]

[a]Department of Psychology, University of California, Berkeley, CA 94720

Edited by Dale Purves, Duke University, Durham, NC, and approved July 10, 2019 (received for review November 21, 2018)

The approximate number system (ANS) has attracted broad interest due to its potential importance in early mathematical development and the fact that it is conserved across species. Models of the ANS and behavioral measures of ANS acuity both assume that quantity estimation is computed rapidly and in parallel across an entire view of the visual scene. We present evidence instead that ANS estimates are largely the product of a serial accumulation mechanism operating across visual fixations. We used an eye-tracker to collect data on participants' visual fixations while they performed quantity-estimation and -discrimination tasks. We were able to predict participants' numerical estimates using their visual fixation data: As the number of dots fixated increased, mean estimates also increased, and estimation error decreased. A detailed model-based analysis shows that fixated dots contribute twice as much as peripheral dots to estimated quantities; people do not "double count" multiply fixated dots; and they do not adjust for the proportion of area in the scene that they have fixated. The accumulation mechanism we propose explains reported effects of display time on estimation and earlier findings of a bias to underestimate quantities.

mathematical cognition | approximate number system | eye-tracking

From infancy, humans are equipped with an approximate number system (ANS) that allows for inexact quantity estimation and comparison (e.g., refs. 1–4). This system is shared with our close and distant evolutionary relatives (e.g., refs. 5–7) and may be related to the development of exact numerical concepts and later mathematics in humans (4, 8–10). However, the defining feature of the ANS is that it is inexact, providing an approximate representation of quantity, which is likely useful in a variety of evolutionary contexts (e.g., refs. 6, 7, 11, and 12). The acuity of an individual's ANS is often quantified in terms of their *Weber fraction*, w, which is a real number denoting how the noise in a representation scales with numerosity. Specifically, one popular psychophysical model of the ANS assumes that a number n is represented by a Gaussian with mean n and SD $w \cdot n$, so that a lower w implies a higher-fidelity system.

The mechanisms supporting the ANS are often contrasted with other mechanisms for computing numerosity, such as counting and subitizing (13–15). Counting, for instance, is dependent on intentional, serial enumeration of a set; the ANS, in contrast, is often viewed as parallel, rapid, and automatic. This view is supported by response times, where counting takes around 300 ms per enumerated item, but approximate-number computations can take as little as 16 ms, independent of the number of objects (16). Additionally, researchers have identified populations of neurons that respond similarly for sequentially and simultaneously presented numerosities in monkeys (17), which has been taken as evidence that ANS representations are not the result of sequential processing.

However, recent evidence has muddied the simple picture of the ANS. Several studies have shown that individuals' Weber fractions are highly task-dependent, differing between estimation and discrimination tasks (e.g., refs. 18 and 19). In fact, Weber fractions have poor retest reliability, even when measured by using the same task (20). Numerical estimates have also been found to be influenced by nonnumerical features of stimuli, such as the degree of clustering in a scene (21). Finally, the precision of numerical estimates is known to improve as stimuli are presented for a longer duration (16), suggesting that ANS estimation may involve some type of temporal process.

Despite this, prior computational models of the ANS have built speed and parallelism into their architecture. For instance, many of the dominant ANS models are feedforward neural network models, where input is processed in parallel and instantaneously (e.g., refs. 22–24). The objective of the present study is to critically evaluate the simple picture of the ANS as a rapid and entirely parallel process. In particular, we aim to capture the possible sequential mechanisms involved in numerical estimation using behavioral experiments and model-driven analysis. We present a model and behavioral data from two experiments that challenge the standard parallel perception theory. Our results lend support instead to an account of ANS estimation that involves sequential integration across visual fixations.

We ran estimation (Experiment 1) and discrimination (Experiment 2) tasks in which participants made nonsymbolic numerosity judgments at different exposure durations. Critically, we collected visual fixation data using an eye-tracker so that we could measure how participants' ANS estimation was influenced by their path of visual fixations. We show that ANS estimates are the result of a serial accumulation process (25), such that estimates increase as a function of foveation. We present an analysis that quantifies the contribution of foveal, peripheral, and multiply fixated dots in an array which supports this interpretation. Our results suggest that individual differences in ANS acuity may

Significance

The question of how people estimate numerical quantities is centrally important in cognitive psychology, neuroscience, and applied educational research. It is generally believed that estimation of numbers is rapid and occurs in parallel across a visual scene. Here, we show that people's estimates are determined by a sequence of visual fixations, with both their mean estimates and their precision increasing as a function of how many points they foveate. This mechanism suggests that a considerable body of research which treats estimation as a purely numerical measure is likely to be missing an important part of the picture: Numerical estimation ability is closely tied to the mechanisms that control eye movements and attention.

Author contributions: S.J.C. and S.T.P. designed research; S.J.C. performed research; S.J.C. analyzed data; and S.J.C. and S.T.P. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: All of the code to run the experiment and all data and analysis used in the paper can be found on S.J.C.'s github page (https://github.com/samcheyette/accumulator_paper_files).

[1]To whom correspondence may be addressed. Email: sjcheyette@gmail.com.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1819956116/-/DCSupplemental.

reflect differences in cognitive processes that are not directly related to numerical estimation, including attention or visual-processing speed. This dependence on nonnumerical factors may explain why studies that train people's ANS yield mixed results in transferring to mathematical knowledge (26–29).

Experiment 1

Since the visual mechanisms supporting the ANS have not been explored in detail, we first used the simplest paradigm possible to understand ANS estimation. Fig. 1 illustrates the sequence of displays shown on each trial. After viewing a fixation cross, participants were shown an array of randomly placed dots on a screen which were noise-masked after a short time. They were then prompted to enter an estimate in Arabic numerals. Subjects were not given feedback and thus had no push to recalibrate their response scale. We manipulated the amount of time that each display was visible to replicate and extend the previous work on the effect of timing on the ANS. Each participant completed 16 trials for each of four display-time conditions: 100; 333; 1,000; and 3,000 ms. We eye-tracked participants during this task to determine how their estimates were influenced by the number of dots in the path of their visual fixation. Importantly, the screen subtended a range of participants' visual field that ensured that some of the dots could be seen foveally and others peripherally from the initial fixation.[*]

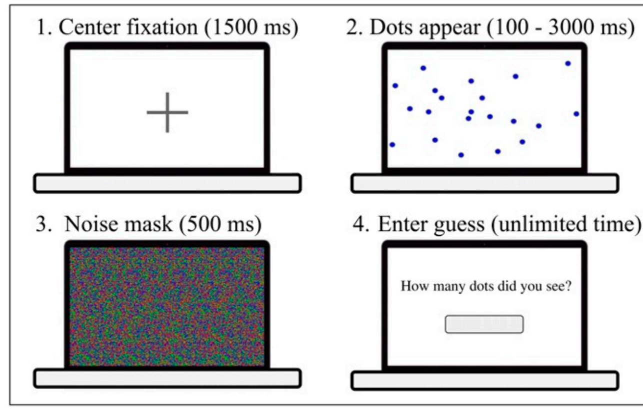

Fig. 1. Each of the four images represents one stage of a trial in the estimation task, in their order. Stage 1: A fixation cross appears for 1,500 ms. Stage 2: The fixation cross is removed, and dots appear on the screen for between 100 ms and 3 s, depending on the condition. Stage 3: The display is masked by noise for 500 ms. Stage 4: A prompt appears asking for an estimate of the number of dots shown.

Results.

Replication of basic number psychophysics. Fig. 2*A* shows how the mean estimate (*y* axis) varied as a function of the quantity displayed (*x* axis), collapsing over all time conditions. There are two aspects of this graph worth highlighting: First, mean estimates vary approximately linearly as a function of quantity, exactly as should be found in Weber models of the number system. Second, this shows a strong tendency to increasingly underestimate larger numbers,[†] as shown by the fact that the slope of the line is less than 1, which would have corresponded to perfectly veridical estimation (assuming an intercept of 0). Both effects have been found robustly in the literature (e.g., ref. 31). Fig. 2*B* shows that Experiment 1 replicates the second traditional property of ANS estimation: *scale variability*, wherein the error in estimation increases linearly in magnitude.

More time improves estimation mean and variance. To evaluate whether timing influenced participants' ANS, we ran a hierarchical regression to estimate the effect of time on both the mean estimate and Weber fraction, including participant- and group-level regression effects fit jointly. The model assumed that means and SDs varied linearly as a function of quantity in accordance with Weber's law. More specifically, on a trial that showed n dots, each participant's mean estimate was drawn from a Gaussian centered around $\beta \cdot n$ and modeled with SD $w \cdot \beta \cdot n$, where β and w are hierarchically fit parameters (*SI Appendix*). The regression included logarithmic effects of time on mean estimates and Weber ratios, allowing us to extract each individual's effective slope and Weber ratio as a function of time.

Fig. 2 shows the mean slope (Fig. 2*C*) and Weber fraction (Fig. 2*D*) in each time condition extracted from this model. The group-level means are shown in blue, and each participant is shown by a line in black. If participants' estimates were unbiased (e.g., veridical estimation as opposed to underestimation), then the group mean slopes would be 1 (black dotted line), and if time did not have an effect, the group mean slopes and Weber fractions (y axis) would remain constant across time (x axis). In contrast, Fig. 2*C* shows that subjects consistently underestimate with slopes less than 1, but that this underestimation effect decreases with increasing time. Participants' average slope increases by about 17% (0.71–0.83) from the shortest to the longest time condition. This is what would be expected by quantity accumulation over time: More time increases reported quantities. Additionally, their average Weber fraction decreases by about 21% (0.28–0.22) (*SI Appendix*, Table S1). Correspondingly, Fig. 2*D* shows that Weber fractions improve (decrease) with more time.

Foveation, not time, is what matters for estimation. If ANS estimation is driven by accumulation of quantity across saccades, we should first expect that mean estimates increase with foveation. We should also expect that time has no effect when jointly considering foveation—i.e., that time simply allows for more saccades and nothing more. To evaluate this, we summed the number of dots that fell within 5° (often called the "parafoveal region") of the center of participants' fixation paths for more than 50 ms on a trial.[‡] We denote the dots that are seen for at least this amount of time as "foveated."

Fig. 3 provides four example trials, depicting a participant's gaze path across the screen while the stimulus is being shown. The filled points represent "foveated" dots, and the unfilled points represent those that were not.[§] At the bottom of each display, the number of dots shown, foveated, and estimated are shown. We provide a more rigorous formalization and test of this idea in The Mechanics of ANS Estimation.

Fig. 4*A* shows the percent of dots that are foveated for each time condition. As should be expected, more dots are foveated with longer exposure duration. The average proportion of dots foveated more than tripled from the shortest to longest time condition (18–64%). Consistent with the hypothesis that effects of time are due to accumulation of foveated dots, the effects of time on estimation disappeared when the effect proportion of dots foveated was jointly taken into account. Fig. 4*B* shows the percent deviation of estimates from the true quantity as a function of dots foveated, colored by time. That the lines overlap suggests

[*] The display size was 38° of participants' visual field left-to-right and 26° top-to-bottom. This is smaller than some displays used in prior ANS literature (e.g., ref. 30), but large enough that some dots are peripheral.

[†] When we use the term "underestimation," we mean that the average estimate is less than a given numerosity.

[‡] We also tested 16 ms and 100 ms as possible thresholds and 2° and 10° as possible visual degree thresholds. These differences did not affect the qualitative pattern of results.

[§] This is for illustrative purposes only—stimuli were entirely static during a trial.

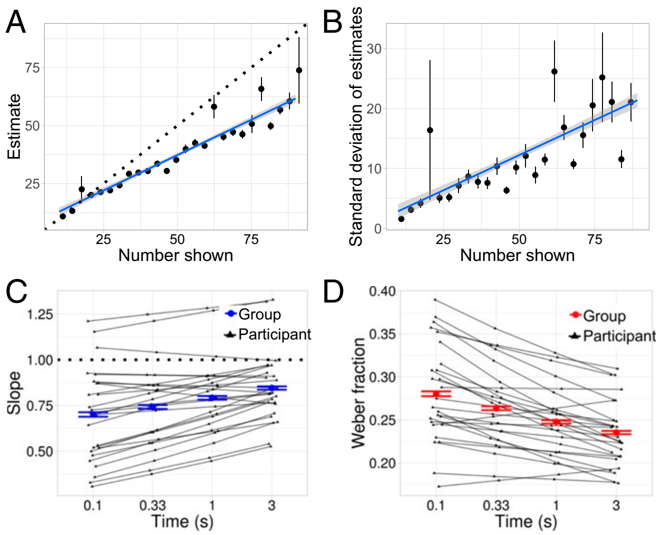

Fig. 2. (*A*) Estimates as a function of the number of dots presented, collapsing across time conditions. Points are binned means, with errors representing bootstrapped 95% CIs. (*B*) The SD of participants' estimates as a function of the number of dots displayed, collapsing across time conditions. (*C*) Participant (black) and group-level (blue) slopes in each time condition of the estimation task are shown. Slopes represent the way the mean estimate scales as a function of quantity shown. (*D*) Participant (black) and group-level (red) Weber fractions in each time condition of the estimation task are shown.

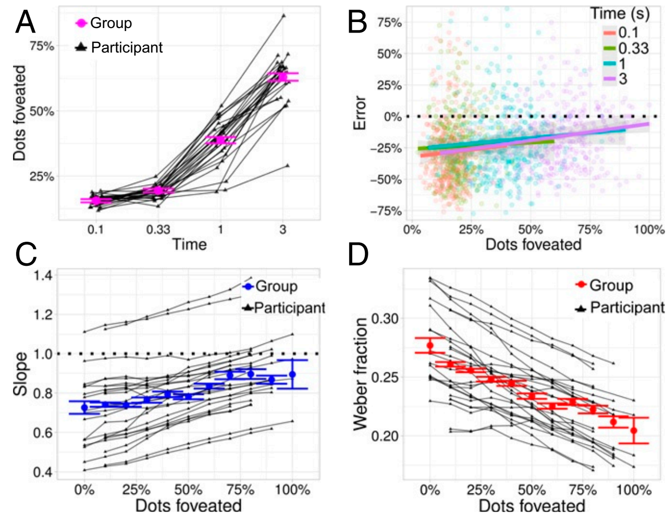

Fig. 4. (*A*) The proportion of dots foveated (*y* axis) as a function of time (*x* axis), at the group level (red) and for each participant (black). (*B*) The percent deviation of estimates from the true number of dots (*y* axis) as a function of the percent of dots foveated (*x* axis). Each time condition is grouped by color. (*C*) The slope of participants' mean estimates (*y* axis) as a function of the percent of dots foveated (*x* axis). (*D*) Weber fractions (*y* axis) as a function of the percent of dots foveated.

that there is no effect of time when both foveation and time are taken into account.

To formally evaluate this, we ran a second hierarchical regression that was identical to the one reported above, except that it included a covariate for the effect of the proportion of dots foveated on the mean and variance of each participant's estimate. This regression shows that the proportion of dots foveated significantly affected both the mean of participants' estimates (Fig. 4*C*) and Weber ratios (Fig. 4*D*), and the effect of time disappeared when foveation was taken into account. Moreover, when 100% of dots were foveated, participants were nearly unbiased (slope ≈ 1 in Fig. 4*C*), suggesting that the underestimation bias previously observed was not miscalibration, but was, rather, due to participants not foveating all of the dots. In a separate analysis, we found that the observed effect of foveation on mean estimates held in each time condition separately (*SI Appendix*, Table 3).

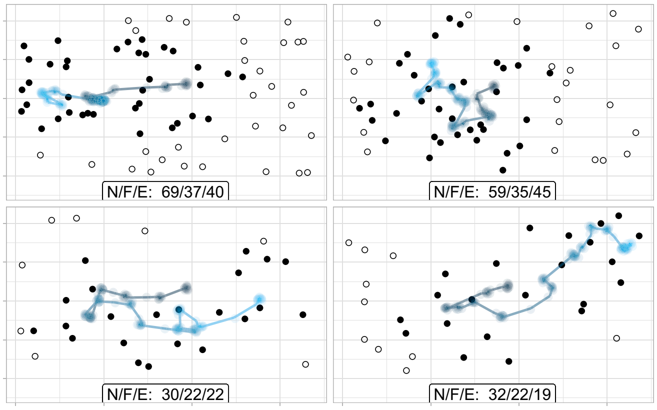

Fig. 3. Example fixation paths of one subject in the 3-s time condition, with each panel representing a single trial. The points represent the dots displayed on their screen, where filled dots represent the ones that were foveated. At the bottom of each panel, a label N/F/E shows how many dots were shown (*N*), how many were foveated (*F*), and what quantity the participant actually estimated (*E*).

Thus, these results provide an alternative account of prior findings of *1*) underestimation and *2*) effects of time. Indeed, both are unified into an account where serial accumulation of foveated dots drives numerical quantity estimates. This finding calls into question the construct validity of Weber ratios as a measure of numerical cognition, since numerical estimates depend on how many dots happen to be foveated, a capacity which is nonnumerical.

Experiment 2

Because there is evidence that Weber fractions may differ between estimation and discrimination tasks (19), it is important to replicate these patterns in a discrimination task. We ran a second experiment with the same participants as Experiment 1, again recording participants' gaze. Participants saw two stimuli of dot arrays (as in Fig. 1) sequentially and were then asked to indicate which had a greater quantity. We manipulated timing in four conditions, which determined whether the first or second array of dots was visible for longer. Specifically, we crossed long and short durations to give presentation times of 100:100 ms, 1,000:100 ms, 100:1,000 ms, and 1,000:1,000 ms for the two displays. We predicted that, if ANS estimation relied on foveal accumulation in this task as well, timing would bias participants toward whichever display was presented for longer.

Results. Participants' responses as a function of ratio collapsed across time conditions can be seen in Fig. 5 *A* and *B*. Fig. 5*A* shows the proportion of participants who responded that the second display had more dots than the first as a function of the ratio of dots in the second display relative to the first. The proportion of participants who responded that the second display was more numerous increased monotonically with the ratio. Participants reported that the second display was more numerous on average (56% of the time), possibly suggesting an effect of memory. This is consistent with studies finding effects of recency in nonsymbolic magnitude comparison (32). Fig. 5*B* shows participants' accuracy as a function of the absolute magnitude ratio, or the minimum magnitude over the maximum. Participants were able to discriminate ratios of $5:6$ with roughly 75% accuracy.

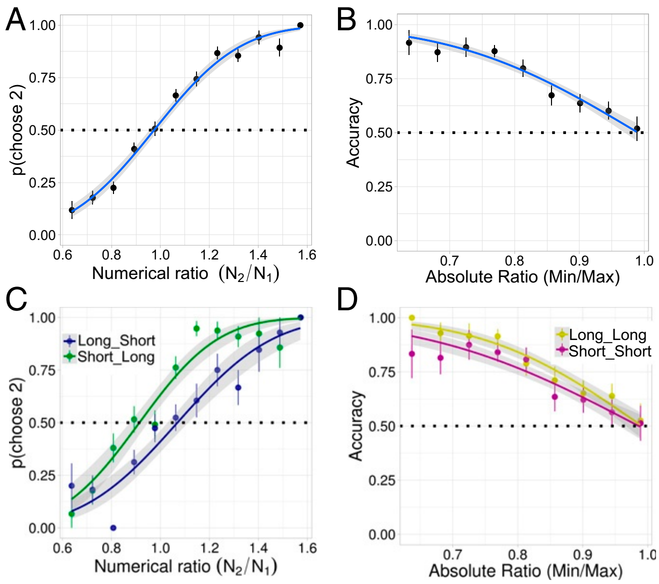

Fig. 5. (*A*) The probability that participants responded that the second display had more dots as a function of the ratio N_1/N_2, where N_1 and N_2 are the number of dots in the first and second display, respectively, collapsed across conditions. The fit curve (as well as all other fits in this display) is from a probit regression. (*B*) Accuracy as a function of the absolute ratio ($Min(N_1, N_2)/Max(N_1, N_2)$). (*C*) The probability participants responded that the second display had more dots in the long–short (blue) and short–long (green) conditions as a function of ratio. (*D*) Accuracy as a function of the absolute ratio in the long–long (yellow) and short–short (red) conditions.

Fig. 5*C* shows response curves for the critical conditions where the first and second displays were shown for different amounts of time, but the total presentation time was controlled (long–short versus short–long). The difference between the curves indicated a bias to choose the second display when it was long compared with when it was short, as predicted. Fig. 5*D* shows response curves for the conditions where the first and second displays are shown for the same amount of time, but overall presentation time differed (short–short versus long–long). The observed difference between the response curves in Fig. 5*D* indicates that responses in the long–long condition were more accurate than those in the short–short condition. Collapsing across ratios, participants chose the second display 62% of the time in the short–long condition and 45% of the time in the long–short condition, as predicted. Participants chose the second display at intermediate (though above-chance) rates in the short–short (56%) and long–long (57%) conditions (see *SI Appendix* for analysis).

Analogous to the analysis for Experiment 1, we used the eye-tracking data to determine whether participants' visual samples mediated the observed effect of time. We found that the proportion of dots foveated had a significant effect on both the slope and Weber fraction, and, as with the estimation task, the effect of time on the slope disappeared. There was still an effect of time on Weber fraction, though it was heavily reduced (*SI Appendix*).

The Mechanics of ANS Estimation

We next developed a statistical model that allowed us to use people's behavioral data to quantify how different components of visual input contributed to numerical estimates. This model was parameterized in a way that allowed us to test a variety of a priori plausible hypotheses about how ANS accumulation might relate to visual behavior. Primarily, this allowed us to test separable contributions of several behaviorally measured factors to an estimated quantity μ. The weight of each factor was inferred by the model. Fig. 6*A* shows all of these terms in the full equation for μ, with the fit parameters in color and the behaviorally measured variables on each trial in black.

The model assumed that there were five components that contributed to μ. First were the number of dots foveated (N_{foveal}) and the number of dots not foveated ($N_{peripheral}$), which were each weighted by their corresponding regression parameters (β_{foveal} and $\beta_{peripheral}$). In addition, we tested the contribution of dots that were fixated more than once after first saccading away (N_{double} weighted by the parameter β_{double}). Finally, the proportion of area that has been foveated ($A_{foveated}$)—which we measured as percent of the screen within the 5° window used above—and the area not foveated ($A_{peripheral}$) were allowed as scaling factors.

To fit this model to behavioral data, we again used a hierarchical Bayesian model which allowed partial pooling of parameters. Examination of the inferred parameters allowed us to characterize the mechanisms of ANS estimation in three critical ways: First, comparison of $\beta_{peripheral}$ and β_{foveal} will show if the accumulation mechanism relies more, less, or equally on foveal and peripherally observed dots. This, in turn, tells us whether the ANS is primarily parallel or whether foveated dots contribute more to the observed estimates. Second, examination of β_{double} will tell us whether participants "double count" dots that are refoveated ($\beta_{double} \approx 1$) or not ($\beta_{double} \approx 0$). This will answer a basic question about ANS accumulation: Is it based on mere retinal input or on a spatially based picture of the world that is built up across saccades (e.g., ref. 33)? Third, do participants rescale their input by the area they have foveated ($\gamma_{foveated} \approx 1$) to correct for their limited visual sample? Or, is estimation a more simple *accumulator* ($\gamma_{foveated} \approx 0$) that does not take into account how much of the scene has been viewed? Note that our formalization does not test whether area, density, convex-hull, or some other continuous quantity is the basis of numerical estimation (10, 15, 34). Rather, this tests whether the ANS relies preferentially on foveated objects and whether it adjusts for the proportion of screen area that has been foveated. *SI Appendix*, however, presents results showing that convex-hull has little to no effect on mean estimates.

Fig. 6*B* shows the inferred group- and subject-level means for β_{foveal} (*x* axis) and $\beta_{peripheral}$ (*y* axis). This shows that foveated dots contribute about twice as much as peripheral dots to estimates. Moreover, the value of β_{foveal} is $\sim$1, meaning that people veridically count one foveated dot as one more in their estimate.[¶] Interestingly, however, the peripheral dots do provide a nonzero contribution, explaining why ANS estimation is possible with very fast presentation times, albeit with a lower precision (16). Fig. 6*C* shows that both γ_{foveal} and $\gamma_{peripheral}$ are near zero, indicating little area renormalization. This finding supports our primary claim that the estimation is based on accumulation, rather than inference using the density of dots observed in part of the scene. Finally, β_{double} was near 0 for all participants, indicating that there is almost no effect of seeing the same dot multiple times in the same display. This would happen, for instance, if people build up a mental image of the dot array that is fed to the accumulator.

Fig. 6*D* visualizes the relative contribution of each factor to mean estimates (*y* axis) across time conditions (*x* axis), as inferred by the model. At 0.1 s, peripheral and foveated dots contribute roughly equal amounts to estimates, accounting for the significant degree of underestimation given such a short exposure. However, as the exposure time increased, foveated dots contributed increasing amounts to the estimate, such that peripheral dots barely played a role in estimation at 3 s. Rescaling and double-counting played almost no role at any amount of time.

[¶]This does not mean that they were actually counting, as the short display times precluded that. Rather, it means that if all dots in a scene were foveated, estimates would be unbiased in expectation, though not error-free.

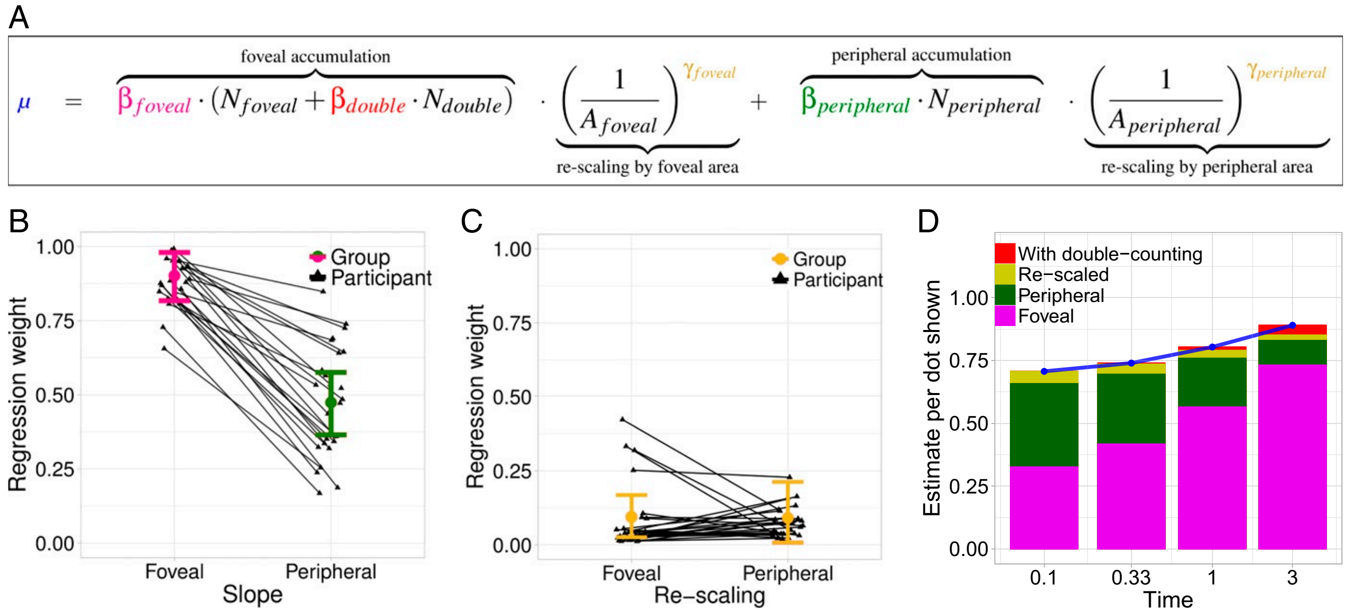

Fig. 6. (*A*) The mean estimate, μ, given as a function of the number of dots foveated, N_{foveal}; the number of dots not foveated, $N_{peripheral}$; the percent of screen area foveated, A_{foveal}; the percent of screen area not foveated, $A_{peripheral}$; and the number of dots foveated more than once, N_{double}. Each of these has a corresponding parameter quantifying its contribution to the estimate μ. (*B*) Parameters β_{foveal} and $\beta_{peripheral}$ capture the foveal and peripheral contribution to the accumulated count. (*C*) Parameters γ_{foveal} and $\gamma_{peripheral}$ capture the degree to which the accumulated count is normalized by the percent of screen for area foveated (A_{foveal}) or peripheral ($A_{peripheral}$). (*D*) A visualization of how each factor contributes to μ over time. As exposure time increases, the average proportion of dots foveated increases, leading to differences in the expected contribution of each factor to the mean estimate.

General Discussion

ANS estimation is typically thought to operate rapidly and in parallel. There is evidence to support this view. For instance, people can discriminate quantities at above-chance levels given only 16 ms of exposure (16). Studies have also demonstrated that reaction times are roughly constant across numerosities in humans and monkeys performing approximate numerical estimation and discrimination tasks (35, 36). The latency of number-sensitive neurons tends to be independent of numerosity in monkeys as well (17, 36). However, the results and analysis we present support an alternative theory: that ANS estimation relies on a serial accumulation mechanism that integrates information—either numerical quantity itself or lower-level visual content—across eye fixations.

Our experiments first replicate two prior behavioral findings: an underestimation bias (31) and a dependence of ANS acuity on time (16). We then showed that the underestimation bias decreases with time, such that participants estimated higher numbers as the stimulus' duration increased. Such an influence of time is predicted by an accumulation model, but not by prior accounts that attribute underestimation to miscalibration of response scales (31). Finally, we showed that the effect of time is almost entirely mediated by visual fixations, suggesting that time matters, because with more time, subjects are able to fixate more of the display. Freely fit parameters from our model indicated that foveated points contribute twice as much to a numerical estimate as peripheral ones. This analysis also revealed that the accumulation likely does not adjust for area, nor does it double-count refixated dots. Together, these results suggest that a primarily foveal, serial accumulation mechanism is at the heart of ANS estimation, rather than the rapid, parallel mechanism previously proposed and commonly imagined.

A serial accumulator is similar to ANS models that perform temporal integration of, for instance, sequences of clicks (5), as well as an approximate version of counting logic observed in sequential presentation of quantities to primates (37). Thus, visual ANS estimation may share resources and processes with nonvisual quantity estimation, as experiments on cross-modal matching would suggest (38). Specifically, visual fixations may be a proxy for attention, which would be consistent with the finding that the numerosity of auditory and tactile stimuli are increasingly underestimated as their presentation rate increases (39, 40). Still, it is surprising to see such serial effects in visual displays since vision can in principle support parallel processes (41).

One limitation of the current work is that our results do not address the specificity of the accumulation mechanism. In particular, our results are consistent with at least two possibilities: Either numerical quantities themselves are being integrated across visual fixations, or people build up an increasingly precise image of the visual scene as they saccade, from which numerical information is later extracted. In either case, our results do show that performance in ANS tasks is largely determined by the serial component of this process.

Regardless of the ultimate mechanisms, our results raise an important methodological point for both basic cognitive research on the ANS and applied education research which relies on it. In light of our findings, it is difficult to interpret results from studies that compare participants' performance across ANS tasks which use different display sizes or stimulus-exposure durations (e.g., refs. 13, 19, and 30). More broadly, our results suggest that the nearly universal use of ANS tasks to index a "pure" sense of number may be misguided. A full picture of ANS estimation will require integrating aspects of visual cognition such as attention and ocular-motor control to understand the cognitive mechanisms that translate visual scenes into abstract numerosities.

Materials and Methods

Experiment 1. Participants were placed directly in front of a computer, with the eye-tracker mounted on top. The computer screen was 24 inches, with an aspect ratio of 16:10 and screen resolution of 1,920 × 1,200 pixels. The screen sat on an adjustable desk, which was vertically realigned for each participant to ensure that that the center of the screen was level with their eyes.

The participants were fixed to a distance such that their eyes were 26 inches away from the screen, which was ensured by measurement with a yardstick. The screen subtended ∼38° of participants' visual field left-to-right and 26° top-to-bottom. The eye-tracker was a Tobii T60XL model, providing a readout of 60 samples per second. We used built-in Tobii software to calibrate participants to the eye-tracker.

Each dot had a radius of 10 pixels. The density of the dots in the images ranged from 0.01 to 0.07 dots/deg^2. Note that the constant dot size meant that it was not possible to directly test whether the ANS estimation uses number rather than another correlated dimension such as density. The number of dots displayed on each trial varied between 10 and 90 dots, inclusive. To determine the numerosities shown to a given subject, 16 numbers were chosen randomly from within that range. The same 16 numbers were shown to the participant in each of the 4 time conditions, with presentation order randomized across the conditions. The median range size across participants was 71 (minimum 54, maximum 79). The median lowest number shown was 14, and the median highest number shown was 86. The dots were placed on the screen at random locations, only constrained to be nonoverlapping. Participants entered their numerical estimates using a keyboard attached to the computer. The experiment was designed by using the Python library Kelpy (42); all of the code to run it can be found on the first author's github page at https://github.com/samcheyette/accumulator_paper_files. All data and analysis used in the paper can be found there as well.

Participants. A total of 27 adult subjects (15 female, 12 male) from the University of Rochester community were recruited to participate in the task. The participants' ages ranged from 18 to 29 ($M = 21.4$).

Procedure. All study procedures were approved by the University of Rochester Institutional Review Board. After providing consent, participants were calibrated to the eye-tracker and subsequently began the experiment. The experiment consisted of 64 total trials, with 4 blocks of 16 trials each. Each 16-trial block contained one of the 4 different time conditions each subject underwent: 100; 333; 1,000; and 3,000 ms (together comprising all 64 trials); the order of the blocks was randomized across participants. On each trial, dots were displayed, followed by a noise mask. Subjects then typed their responses into a text box using a keyboard and pressed the enter key to move onto the next trial.

Experiment 2. Experiment 2 was a sequential number-discrimination task with the same participants who completed Experiment 1. Likewise, the properties of the stimuli and materials used in Experiment 2 were the same as in Experiment 1 (e.g., dots in both had the same radius). Sixteen pairs of numbers were chosen randomly for each participant, with the ratio (the minimum over the maximum) of the number pairs constrained to be between 0.5 and 0.99. For a given subject, the same number pairs were used across the 4 time conditions, with their order randomized across conditions.

Procedure. After completing Experiment 1, participants took a break (if needed), were recalibrated to the eye-tracker, and then began Experiment 2. In this task, participants saw 2 flashes of dots, one after the other, and were subsequently asked which stimulus they thought had a greater quantity of dots (pressing 1 or 2 on a keyboard). There were 4 conditions with 16 trials each (like Experiment 1), where each condition corresponded to a unique pair of stimulus durations for the first and second display. More specifically, the 4 conditions were (100; 100 ms), (100; 1,000 ms), (1,000; 100 ms), and (1,000; 1,000 ms).

ACKNOWLEDGMENTS. We thank Ashley Barhdan for her help running subjects; Celeste Kidd, Jessica Cantlon, Dick Aslin, Shirlene Wade, Willa Voorhies, and Fred Callaway for useful feedback and suggestions; and the Kidd laboratory for use of their eye-tracker. This work was supported by National Science Foundation, Division of Research on Learning Grant 1760874 (to S.T.P.) and *Eunice Kennedy Shriver* National Institute of Child Health and Human Development at the National Institutes of Health Award 1R01HD085996 (to S.T.P. and Jessica Cantlon).

1. F. Xu, E. S. Spelke, Large number discrimination in 6-month-old infants. *Cognition* **74**, B1–B11 (2000).
2. K. McCrink, K. Wynn, Ratio abstraction by 6-month-old infants. *Psychol. Sci.* **18**, 740–745 (2007).
3. D. C. Hyde, D. A. Boas, C. Blair, S. Carey, Near-infrared spectroscopy shows right parietal specialization for number in pre-verbal infants. *Neuroimage* **53**, 647–652 (2010).
4. S. Dehaene, *The Number Sense: How the Mind Creates Mathematics* (Oxford Univ Press, New York, NY, 2011).
5. W. H. Meck, R. M. Church, A mode control model of counting and timing processes. *J. Exp. Psychol. Anim. Behav. Process.* **9**, 320–334 (1983).
6. T.-I. Yang, C.-C. Chiao, Number sense and state-dependent valuation in cuttlefish. *Proc. R. Soc. B* **283**, 20161379 (2016).
7. J. F. Cantlon, Math, monkeys, and the developing brain. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 10725–10732 (2012).
8. L. Feigenson, S. Dehaene, E. Spelke, Core systems of number. *Trends Cognit. Sci.* **8**, 307–314 (2004).
9. J. Halberda, M. M. M. Mazzocco, L. Feigenson, Individual differences in non-verbal number acuity correlate with maths achievement. *Nature* **455**, 665–668 (2008).
10. A. Starr, N. K. DeWind, E. M. Brannon, The contributions of numerical acuity and non-numerical stimulus features to the development of the number sense and symbolic math achievement. *Cognition* **168**, 222–233 (2017).
11. S. T. Piantadosi, J. F. Cantlon, True numerical cognition in the wild. *Psychol. Sci.* **28**, 462–469 (2017).
12. H. J. Gross *et al.*, Number-based visual generalisation in the honeybee. *PLoS One* **4**, e4263 (2009).
13. S. K. Revkin, M. Piazza, V. Izard, L. Cohen, S. Dehaene, Does subitizing reflect numerical estimation? *Psychol. Sci.* **19**, 607–614 (2008).
14. D. C. Burr, M. Turi, G. Anobile, Subitizing but not estimation of numerosity requires attentional resources. *J. Vision* **10**, 20–20 (2010).
15. G. Anobile, G. M. Cicchini, D. C. Burr, Separate mechanisms for perception of numerosity and density. *Psychol. Sci.* **25**, 265–270 (2014).
16. M. Inglis, C. Gilmore, Sampling from the mental number line: How are approximate number system representations formed? *Cognition* **129**, 63–69 (2013).
17. A. Nieder, I. Diester, O. Tudusciuc, Temporal and spatial enumeration processes in the primate parietal cortex. *Science* **313**, 1431–1435 (2006).
18. G. R. Price, D. Palmer, C. Battista, D. Ansari, Nonsymbolic numerical magnitude comparison: Reliability and validity of different task variants and outcome measures, and their relationship to arithmetic achievement in adults. *Acta Psychol.* **140**, 50–57 (2012).
19. M. Guillaume, W. Gevers, Assessing the approximate number system: No relation between numerical comparison and estimation tasks. *Psychol. Res.* **80**, 248–258 (2016).
20. M. Inglis, C. Gilmore, Indexing the approximate number system. *Acta Psychol.* **145**, 147–155 (2014).
21. H. Yeon Im, S.-h. Zhong, J. Halberda, Grouping by proximity and the visual impression of approximate number in random dot arrays. *Vision Res.* **126**, 291–307 (2016).
22. S. Dehaene, J.-P. Changeux, Development of elementary numerical abilities: A neuronal model. *J. Cognit. Neurosci.* **5**, 390–407 (1993).
23. T. Verguts, W. Fias, Representation of number in animals and humans: A neural model. *J. Cognit. Neurosci.* **16**, 1493–1504 (2004).
24. I. Stoianov, M. Zorzi, Emergence of a 'visual number sense' in hierarchical generative models. *Nat. Neurosci.* **15**, 194–196 (2012).
25. C. R. Gallistel, R. Gelman, Non-verbal numerical cognition: From reals to integers. *Trends Cognit. Sci.* **4**, 59–65 (2000).
26. J. Park, E. M. Brannon, Training the approximate number system improves math proficiency. *Psychol. Sci.* **24**, 2013–2019 (2013).
27. D. C. Hyde, S. Khanum, E. S. Spelke, Brief non-symbolic, approximate number practice enhances subsequent exact symbolic arithmetic in children. *Cognition* **131**, 92–107 (2014).
28. J. Park, V. Bermudez, R. C. Roberts, E. M. Brannon, Non-symbolic approximate arithmetic training improves math performance in preschoolers. *J. Exp. Child Psychol.* **152**, 278–293 (2016).
29. M. Lindskog, A. Winman, No evidence of learning in non-symbolic numerical tasks—a comment on Park and Brannon (2014). *Cognition* **150**, 243–247 (2016).
30. J. Halberda, L. Feigenson, Developmental change in the acuity of the "number sense": The approximate number system in 3-, 4-, 5-, and 6-year-olds and adults. *Dev. Psychol.* **44**, 1457–1465 (2008).
31. V. Izard, S. Dehaene, Calibrating the mental number line. *Cognition* **106**, 1221–1247 (2008).
32. R. Van den Berg, M. Lindskog, L. Poom, A. Winman, Recent is more: A negative time-order effect in nonsymbolic numerical judgment. *J. Exp. Psychol. Hum. Percept. Perform.* **43**, 1084–1097 (2017).
33. M. J. Farah, K. M. Hammond, D. N. Levine, R. Calvanio, Visual and spatial mental imagery: Dissociable systems of representation. *Cognit. Psychol.* **20**, 439–462 (1988).
34. D. Odic, J. Halberda, Eye movements reveal distinct encoding patterns for number and cumulative surface area in random dot arrays. *J. Vision* **15**, 5–5 (2015).
35. G. Mandler, B. J. Shebo, Subitizing: An analysis of its component processes. *J. Exp. Psychol. Gen.* **111**, 1–22 (1982).
36. A. Nieder, D. J. Freedman, E. K. Miller, Representation of the quantity of visual items in the primate prefrontal cortex. *Science* **297**, 1708–1711 (2002).
37. J. F. Cantlon, S. T. Piantadosi, S. Ferrigno, K. D. Hughes, A. M. Barnard, The origins of counting algorithms. *Psychol. Sci.* **26**, 853–865 (2015).
38. P. Starkey, E. S. Spelke, R. Gelman, Numerical abstraction by human infants. *Cognition* **36**, 97–127 (1990).
39. D. M. Forsyth, A. Chapanis, Counting repeated light flashes as a function of their number, their rate of presentation, and retinal location stimulated. *J. Exp. Psychol.* **56**, 385–391 (1958).
40. E. C. Lechelt, Temporal numerosity discrimination: Intermodal comparisons revisited. *Br. J. Psychol.* **66**, 101–108 (1975).
41. A. M. Treisman, G. Gelade, A feature-integration theory of attention. *Cognit. Psychol.* **12**, 97–136 (1980).
42. S. T. Piantadosi, Kelpy: A free library for child experimentation in Python, https://github.com/piantado/kelpy/. Accessed 1 April 2018. (2012).