Psychological Review

Limited Information-Processing Capacity in Vision Explains Number Psychophysics

Samuel J. Cheyette, Shengyi Wu, and Steven T. Piantadosi

Online First Publication, April 22, 2024. https://dx.doi.org/10.1037/rev0000478

AMERICAN PSYCHOLOGICAL ASSOCIATION

Psychological Review

ISSN: 0033-295X

https://doi.org/10.1037/rev0000478

Limited Information-Processing Capacity in Vision Explains Number Psychophysics

Samuel J. Cheyette[1], Shengyi Wu[1], and Steven T. Piantadosi[2]
[1] Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology
[2] Department of Psychology, University of California, Berkeley

Humans and other animals are able to perceive and represent a number of objects present in a scene, a core cognitive ability thought to underlie the development of mathematics. However, the perceptual mechanisms that underpin this capacity remain poorly understood. Here, we show that our visual sense of number derives from a visual system designed to efficiently encode the location of objects in scenes. Using a mathematical model, we demonstrate that an efficient but information-limited encoding of objects' locations can explain many key aspects of number psychophysics, including subitizing, Weber's law, underestimation, and effects of exposure time. In two experiments ($N = 100$ each), we find that this model of visual encoding captures human performance in both a change-localization task and a number estimation task. In a third experiment ($N = 100$), we find that individual differences in change-localization performance are highly predictive of differences in number estimation, both in terms of overall performance and inferred model parameters, with participants having numerically indistinguishable inferred information capacities across tasks. Our results therefore indicate that key psychophysical features of numerical cognition do not arise from separate modules or capacities specific to number, but rather as by-products of lower level constraints on perception.

Keywords: visual perception, visuospatial memory, numerosity perception, information theory, modeling

Supplemental materials: https://doi.org/10.1037/rev0000478.supp

Numerosity perception has been studied for at least 150 years (Jevons, 1871), and its psychophysics has been well characterized by experimental work (e.g., Dehaene, 2011; Feigenson et al., 2004; Jevons, 1871; Kaufman et al., 1949; Revkin et al., 2008). However, a basic unresolved question is whether the behavioral patterns found in estimation result from numerical processing itself or from the perceptual processes that feed into numerical perception. In the first case, people may possess a "number system" that itself is the origin of the phenomena seen in behavioral tasks involving number, such as Weber's law and underestimation. For instance, the noise and bias observed in numerical estimation might arise from a sampling process in which numerical information is extracted from visual representations, rather than from noise inherent to visual representations themselves (Dehaene & Changeux, 1993; Heng et al., 2020; Woodford, 2020). Alternatively, such phenomena may emerge as a consequence of more general visual processes, which precede numerical estimation and indeed feed into it (Anobile et al., 2020; Stoianov & Zorzi, 2012; Testolin, Dolfi, et al., 2020; Trick & Enns, 1997; Zorzi & Testolin, 2018). Under this latter hypothesis, the psychophysics of number in vision could result from constraints inherent to visuospatial memory, and then we would expect people's behavior in nonnumerical visuospatial tasks to show equivalent hallmarks to those seen in estimation. Suggestive of this possibility, a host of studies have found that perception of numerosity—both in the estimation and subitizing ranges—is strongly influenced by purely visual factors such as item arrangement (Anobile et al., 2020; Atkinson et al., 1976; Ginsburg, 1976; Krajcsi et al., 2013; G. S. Starkey & McCandliss, 2014; Trick & Enns, 1997; Yang & Chiao,

Samuel J. Cheyette https://orcid.org/0000-0002-7052-5369

This work was supported by Grants 1901262 and 2000759 from the National Science Foundation, Division of Research on Learning (to Steven T. Piantadosi). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the article. The authors are grateful to Fred Callaway for helpful discussions and Ben Pitt for useful comments on an earlier draft of this article. The authors have no conflicts of interest to disclose.

Experiments 1 and 2 were preregistered; Experiment 3 was run at the suggestion of a reviewer and was not preregistered. The preregistration, anonymized data from all experiments, and modeling code are available at https://osf.io/vgm65/. An early version of this work appeared in the Proceedings of the Cognitive Science Conference in 2021; a preprint of this article was also made available on PsyArXiv.

Samuel J. Cheyette played a lead role in conceptualization, data curation, formal analysis, investigation, methodology, visualization, writing–original draft, and writing–review and editing and an equal role in project administration and software. Shengyi Wu played a supporting role in investigation, methodology, project administration, and writing–review and editing and an equal role in software. Steven T. Piantadosi played a lead role in funding acquisition and supervision, a supporting role in conceptualization, formal analysis, investigation, methodology, writing–original draft, and writing–review and editing, and an equal role in resources.

Correspondence concerning this article should be addressed to Samuel J. Cheyette, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, 43 Vassar Street, Building 46-4053, Cambridge, MA 02139, United States. Email: cheyette@mit.edu

2016), eccentricity (Palomares et al., 2011), presentation time (Cheyette & Piantadosi, 2019, 2020; Inglis & Gilmore, 2013), color contrast (Cheyette & Piantadosi, 2020), and the entropy of nonnumeric features (DeWind et al., 2020; Qu et al., 2022).

Our goal in this article is to formalize and test the relationship between visuospatial memory and the perception of quantity to determine whether the properties observed in the number literature—including exact perception of small quantities, scalar variability for larger quantities, and sensitivity to presentation time—result from more general mechanisms of the visual system. If a model of basic visuospatial processing fit to a nonnumerical task recovers the hallmarks observed in visual number psychophysics, this would suggest that features of number perception should really be considered artifacts of basic vision rather than the number itself. Conversely, if features of numerical perception are not latent in a nonnumerical visual memory task that uses the same stimuli, they have to be the result of specifically numerical processes.

We develop a computational model of bandwidth-limited scene memory, which forms beliefs about where individual objects exist in space; these beliefs can then be straightforwardly converted into beliefs about the number of objects in that scene (i.e., by summing over beliefs about where objects are in the scene). This approach builds on recent neural network models that exhibit some numerical properties as a consequence of imperfectly representing a scene (Kim et al., 2021; Stoianov & Zorzi, 2012; Testolin, Dolfi, et al., 2020; Testolin, Zou, et al., 2020; Zorzi & Testolin, 2018), but here we derive the optimal form of this representation and empirically test the predictions of the optimal model. We show that even though the model is explicitly optimized only to detect and remember the presence of objects in various locations, the resulting probability distributions over numerosities closely match known properties of number psychophysics, including both subitizing and Weber's law. Notably, although the model represents a probability distribution over discrete individuals, it behaves like an "analog magnitude system" (Carey, 2009; Gebuis & Reynvoet, 2012; Lourenco & Longo, 2010) when its information capacity is exceeded.

Our account is closely related to the fingers of instantiation model of spatial indexing developed by Pylyshyn (1989), and later expanded by Trick and Pylyshyn (1993, 1994), who proposed that subitizing is the result of a limited-capacity, parallel process for individuating objects in space. On their account, there is a preattentive visual mechanism for selecting items in space, akin to "pointers," which automatically picks out items' locations in space before feature-binding occurs; the limited number of available pointers for indexing corresponds to the subitizing limit. However, their stated goal was to distinguish subitizing from *counting* to explain response time differences in exact enumeration (Mandler & Shebo, 1982). They did not have an explanation for how people inexactly, but still extremely quickly (Inglis & Gilmore, 2013), approximately quantify larger sets. Our own proposal can be thought of as a formalization and extension of their fingers of instantiation model to cover both small and large cardinalities. However, rather than assuming that individuation works with a fixed and discrete set of pointers, we assume instead that there is a limited amount of information available in early vision for individuating and tracking items' locations in space that can be expended as a continuous resource (Alvarez & Franconeri, 2007; Ma et al., 2014; Van den Berg et al., 2012; Vul et al., 2009). There is thus a fluid transition from the regime of exact representation of items' locations when individuating small sets to approximate representation of items' locations when individuating larger sets, and the model recovers subitizing and inexact estimation (Weber's law) as emergent properties of the same system below and above a capacity bound.

This approach is also an extension and reconceptualization of our previous work showing that a single system can account for the discontinuity in estimation ability at four (Cheyette & Piantadosi, 2020). The key idea in that article was that an efficient encoding of number, using at maximum some number of bits of information, will prioritize representations of small numbers at the expense of large numbers because people tend to need to represent small numbers more frequently (Dehaene & Mehler, 1992; Piantadosi & Cantlon, 2017). That work therefore derived exactness for small numbers (e.g., subitizing) and approximation for large numbers by solving a single, unifying optimization. However, it did not explain the key step of how numerosities are actually computed from visual input. In fact, this model assumed that, all else being equal, small and large numerosities are equally easy to perceive—their differing behavioral signatures being solely a matter of frequency of use. Here, we use a similar formal approach to Cheyette and Piantadosi (2020) but apply it to the visual processing that precedes numerical representations rather than to the number itself.

The resulting model makes predictions about the psychophysics of spatial memory and numerosity perception, and how they should covary as a function of an information-processing capacity limit in visual memory. The model predicts that people's ability to remember the locations of objects in space will be near-perfect for small groups but will degrade proportionally with the number of objects in the scene for larger groups. The model additionally predicts that when there is less available visual information (e.g., given shorter exposure time), people will become increasingly unable to precisely remember the locations of even small groups of objects. We can likewise derive predictions about the psychophysics of numerical estimation as a function of cardinality and information capacity, when the output of this bandwidth-limited system is used as input for numerosity estimation. Many of these predictions have been observed previously. Specifically, the visual model predicts the following: (a) exact or near-exact estimation of small sets (e.g., Jevons, 1871; Kaufman et al., 1949; Mandler & Shebo, 1982); (b) a subitizing range that increases as a function of display time (Cheyette & Piantadosi, 2020; Haladjian & Pylyshyn, 2011; Mandler & Shebo, 1982); (c) roughly normally shaped response distributions for estimation (e.g., Nieder & Dehaene, 2009; Pica et al., 2004); (d) increasingly noisy estimation (scalar variability) for larger sets (e.g., Dehaene, 2011; Feigenson et al., 2004; Jevons, 1871; Nieder & Miller, 2004; Xu & Spelke, 2000); (e) estimation acuity that increases with display time (Alonso-Diaz et al., 2018; Inglis & Gilmore, 2013); and (f) an underestimation bias for large sets (Izard & Dehaene, 2008; Mandler & Shebo, 1982) that diminishes with increased display time (Cheyette & Piantadosi, 2019).

To test these predictions about visuospatial memory and numerical perception, we ran two preregistered experiments[1]: a change-localization task to probe participants' memory for the locations of objects (Experiment 1) and a numerical estimation task (Experiment 2). In both experiments, we manipulated the amount of

[1] The preregistration of the model and both experiments can be found at https://osf.io/vgm65/.

information available to participants by varying the exposure time of the presented objects. We found that participants' ability to remember the locations of objects—both for different exposure times and for different numbers of objects present—predicts the observed psychophysics of number under analogous conditions. A replication of these experiments using a within-subjects design (Experiment 3) found that individual differences in performance on the change-localization task were strongly predictive of individual differences in number estimation, with high correlations in model parameters inferred separately from participants' change-localization and numeric estimation data. Since the patterns of bias and noise observed in numerical estimation can be predicted from the uncertainty people have about items' locations in space, this indicates that the psychophysics of number are governed by limits of visuospatial memory, rather than processing specific to quantity itself.

Model

The model aims to capture how an idealized, information-limited perceptual system would perform assuming its goal is to accurately represent the locations of objects in a scene. Although this formalizes the idea of object memory—not specifically numerical estimation—it implicitly encodes cardinality since its output is beliefs about the objects present in a scene. For a given, observed scene containing objects s, we will consider the probability distribution $Q(s'|s)$, giving the system's belief that s' was observed instead of s. We analytically derive an optimal form of Q, by specifying three components: (a) a prior distribution representing how likely the model is to encounter a given scene a priori, (b) a loss function representing how good or bad a given representation of the scene is, and (c) an information capacity bound, representing the maximum allowable information processing. These three elements define a constrained optimization problem, which can be solved to determine the optimal psychophysical distribution $Q(\cdot|s)$, corresponding to the optimal perceptual system. This method is not identical with, but is somewhat analogous to, Bayesian inference that begins with a prior distribution and combines it with evidence to produce a "posterior" distribution; the key difference is that the shape of the "posterior" $Q(\cdot|s)$ is not derived from Bayes rule, but rather from minimizing the loss function (a) subject to an information bound (b).

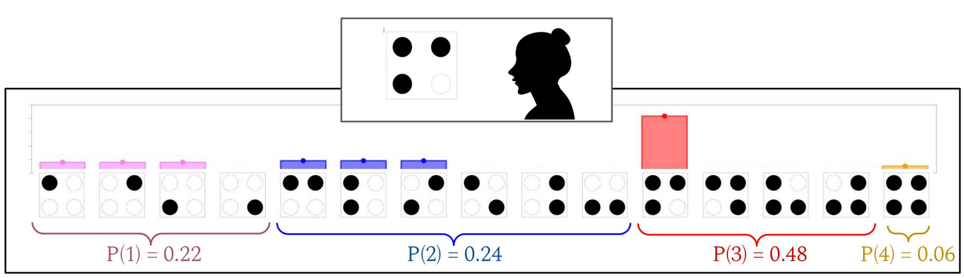
Figure 1 illustrates the basic setup, assuming for the sake of clarity that there are only four possible object locations (or pixels). When a person sees a particular scene, they encode a probability distribution over each possible arrangement of objects, which is a weighted combination of a prior for small numbers and how well the representation matches their observation (akin to a likelihood). This probability distribution can, in turn, be converted into a probability distribution over numerosities by summing the probabilities of each scene with a given number of objects. One key simplifying assumption we make in modeling this setup is that spatial memory encodes the presence or absence of objects in various locations as a discrete matrix. In other words, we assume that visuospatial memory represents a matrix with M black and white pixels to specify where there are objects (and where there are not). We further assume a prior on binary matrices where the number of 1's in a matrix matches the naturalistic frequency of a given number. Specifically, the naturalistic frequency of a number n follows a $\frac{1}{n^2}$ law, where n represents cardinality (Dehaene & Mehler, 1992; Piantadosi & Cantlon, 2017). There are $\binom{M}{n}$ matrices with cardinality n, so a given matrix s with cardinality n has prior $P(s) \propto 1 \Big/ \left(n^2 \cdot \binom{M}{n}\right)$.

When shown the matrix s, we assume the model's goal is to represent s with as high fidelity as possible, remembering whether an object was present at each row i and column j, s_{ij}. We define a loss function $L(s, s')$ specifying how closely a matrix s' matches s or how costly it would be to represent s with s'. Representations s' that have a high degree of overlap with s will have lower penalties and those with less overlap will have higher penalties. We will assume that the loss function $\mathcal{L}(s, s')$ is proportional to some (perhaps unequal) combination of the proportion of false negatives, $P(s'_{ij} = 0|s_{ij} = 1)$, and the proportion of false positives, $P(s'_{ij} = 1|s_{ij} = 0)$:

$$\mathcal{L}(s, s') = \alpha \cdot P(s'_{ij} = 0|s_{ij} = 1) + (1 - \alpha) \cdot P(s'_{ij} = 1|s_{ij} = 0), \quad (1)$$

with α as a weighting parameter, where $0 \leq \alpha \leq 1$. The reason we separate the contribution of false negatives and false positives here is that it is natural to think that the visual system might care about one more than the other. There are, of course, other plausible loss functions, which in fact give qualitatively similar results (see

Figure 1
Conceptual Illustration of the Model



Note. In this example, a person sees a scene with three objects, which is represented as a probability distribution over all possibilities of what she saw. Possible arrangements of objects are grouped by numerosity, shown as different colors. To get the probability of a numerosity k, the model simply sums the probability of all possible scenes with numerosity k, highlighted at the bottom. See the online article for the color version of this figure.

Supplemental Materials)—though we note that the form of this loss function was preregistered.

Given a loss function and prior, we now seek a function $Q(\cdot|s)$ that minimizes the expected loss between possible inputs s and representations s', corresponding to the "best" representation possible. If the set of possible scenes is S, the expected loss is

$$\mathbb{E}[\mathcal{L}(s,s')] = \sum_{s \in S} P(s) \sum_{s' \in S} Q(s'|s) \cdot \mathcal{L}(s,s'). \quad (2)$$

Unconstrained, the function $Q(\cdot|s)$ that minimizes the expected loss would simply correctly encode the scene,

$$Q(s'|s) = \begin{cases} 1, & \text{if } s' = s \\ 0, & \text{otherwise.} \end{cases} \quad (3)$$

However, cognitive systems are constrained by the amount of information they can process over a given span of time. We incorporate this constraint into the model as a bound on the maximum allowable Kullback–Leibler divergence (KL divergence) between the prior distribution $P(\cdot)$ and resultant distribution $Q(\cdot|s)$ over displays. The KL divergence here represents the amount of information in bits needed to represent the resultant distribution $Q(\cdot|s)$ starting with the distribution $P(\cdot)$, which is equivalent to the total amount of information processing required. Given an information bound B, we then have the constraint on KL divergence from $P(\cdot)$ to $Q(\cdot|s)$, often notated $D_{KL}[Q(\cdot|s)||P(\cdot)]$,

$$\sum_{s' \in R} Q(s'|s) \cdot \log \frac{Q(s'|s)}{P(s')} \leq B \quad \forall s \in R. \quad (4)$$

Using the method of Lagrange multipliers (see Supplemental Materials), we can derive an analytic solution to maximize accuracy while keeping the information processing below the information bound B,

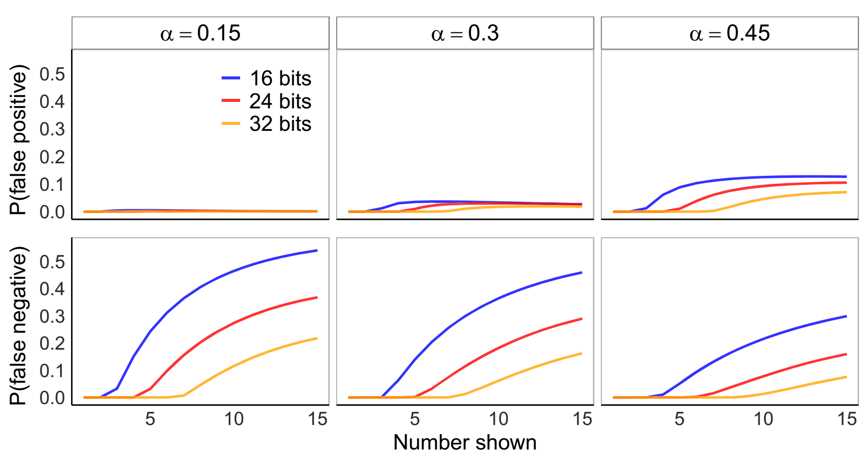
$$Q(s'|s) \propto P(s') \cdot \exp\left(-\frac{P(s)}{\lambda_s} \cdot \mathcal{L}(s,s')\right), \quad (5)$$

for λ_s chosen to satisfy (Equation 4) for each scene s. Note that there is a unique λ_s that satisfies the information-bound constraint for a given scene s and information bound B, which we approximate using numerical methods (see Supplemental Materials for details).

To illustrate how the model works—and the role of the information capacity bound and the loss function parameter α—we generated predictions at different information capacity bounds and with different values of α, assuming a 7 × 7 grid of possible object locations (as will be used in the eventual experiments). The top row of Figure 2 shows the rate at which the model falsely believes an object was in a particular location (a "hallucination") when there was no object there; and the bottom row shows the rate at which the model does not encode an object that was present at a particular location. These predictions are broken down by number (x-axis), information capacity bound (color), and α (columns). The rates of false negatives and false positives increase as a function of quantity at each information capacity bound, reflecting both the decreasing prior over numerosities and the fact that there are more ways to arrange more numerous sets in the range shown. Also apparent is that the model saturates in performance for small quantities when the information bound is high, meaning that it can veridically recall the scene it viewed when there are only a few objects. Finally, this figure illustrates the role of the loss function parameter, α, in controlling the relative cost of hallucinations (top row) versus missing an object (bottom row): as α increases, the ratio of false negatives to false positives increases as well.

Figure 2

Model Predictions for Errors in Spatial Memory



Note. Model predictions for the probability of falsely believing that an object was in a particular location (top row) and falsely believing that there was not an object in a particular location (bottom). The predictions are broken down by quantity (x-axis), information capacity bound (color), and the loss-function parameter α that specifies the cost of false positives relative to false negatives (columns). See the online article for the color version of this figure.

Critically, the model's probability distribution over possible object arrangements s' can be converted into a probability distribution over the total number of objects. Figure 3 shows the *implicit* distribution (y-axis) of numerical estimates (x-axis) for each number 1–15 (lines), at the same information bounds given in Figure 2 (rows). This visual memory model exhibits several key properties of number psychophysics, most notably a transition from exactness to scalar variability. The precise point of transition, as well as the acuity of estimation, is determined by the information bound, as in Cheyette and Piantadosi (2020). We show in Supplemental Materials that the model transitions from subitizing to Weber's law specifically (ratio-dependent discrimination ability). The other variable highlighted in this figure is the loss function parameter α (columns). At low values of α, it is very costly to falsely believe that there was an object present somewhere that there was not (i.e., to hallucinate), so the model predicts very few false positives but has many false negatives—this is true even for small sets under a low information capacity regime. On the other hand, as α increases, it is more costly to miss an object that was present somewhere, which results in less underestimation—and increasingly even overestimation of small quantities.

As Figures 2 and 3 show, there are a range of possibilities for exactly how visuospatial memory and numerical estimation can behave under the model, since the information capacity bound and α are free parameters. Some information capacity bounds would predict veridical estimation up to 20 objects and others would predict no subitizing range at all. Similarly, some values of α predict a bias toward missing objects that were present rather than hallucinating and hence predict underestimating the quantity of observed sets; other values of α predict a bias toward hallucinating rather than missing objects and hence predict overestimating the quantity of observed sets. If this model is correct, then the psychophysics of number are essentially *latent* in visuospatial memory—and probing visuospatial memory in a nonnumeric task should allow us to recover key properties of number representations.

Experiment 1

The goal of Experiment 1 is to understand how visuospatial perception is modulated by processing time and the number of objects in a scene. Our visual model predicts that, given sufficient processing time, participants should be able to remember the locations of small groups of objects with high fidelity but become increasingly inaccurate for larger numerosities, which accords with basic intuition and previous findings (Alvarez & Franconeri, 2007; Vul et al., 2009). With only limited processing time, however, participants should become increasingly incapable of localizing even small groups, and the disparity in performance between smaller and larger groups should decrease, as illustrated in Figure 2. In addition to testing whether localization is well explained by the model, by fitting the information bound and loss parameter α to data gathered from a *non*numerical human spatial memory task, we can test whether the inferred parameters are consistent with the psychophysics of number (Figure 4).

Method

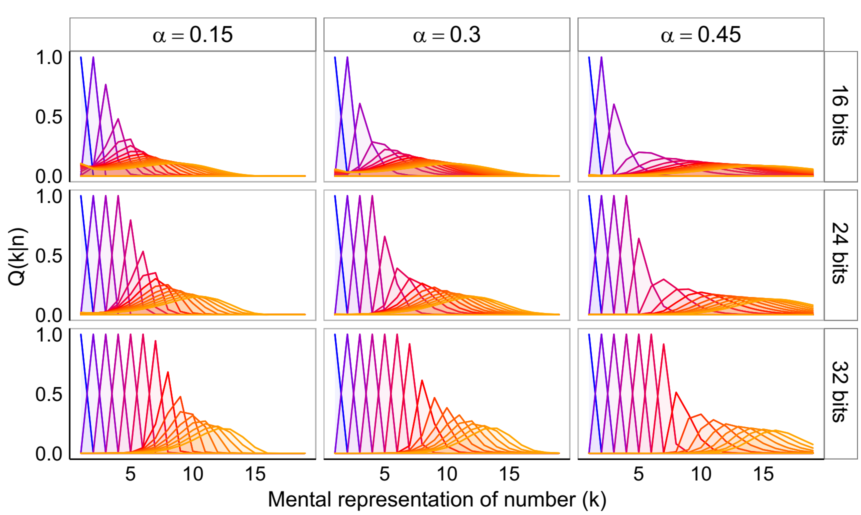
Experiments 1 and 2 were preregistered using the Open Science Framework at https://osf.io/vgm65/. The code for the model and the data collected were also preregistered at https://osf.io/vgm65/. The method, data exclusions, and analyses were all preregistered except where explicitly noted.

Participants

We recruited 110 registered users of Prolific, an online psychology experiment platform. Participants were 18 years old or older, fluent English speakers, and physically present in the United States based on prescreening questions. Each participant who completed the task received compensation of $3. Both experiments were approved by the university's institutional review board and complied with all

Figure 3

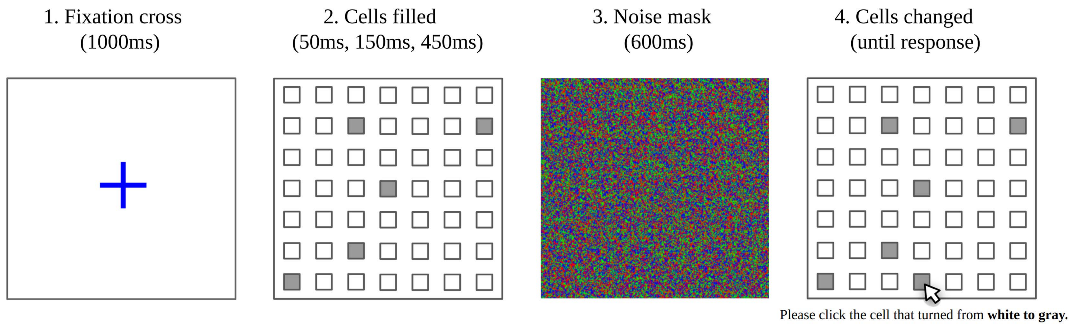
The Implied Number Psychophysics From Spatial Memory



Note. The implied psychophysics of number from the model of spatial memory at different values of α (columns) and information capacity bounds (rows). Each line shows beliefs $(Q(k|n))$ over estimates (k) given numbers $n = 1 \ldots 15$. See the online article for the color version of this figure.

Figure 4

Illustration of Experiment 1



Note. Participants were first shown a fixation cross, followed by a 7 × 7 grid with some of the cells (1–15) filled in gray. A noise mask then appeared after a short time (50 ms, 150 ms, or 450 ms). In the final step, participants were shown a display identical to the one shown previously except for a single cell—one of the previously gray cells either turned white ("disappearance") or one of the previously white cells turned gray ("appearance"). Participants tried to guess which cell had changed. See the online article for the color version of this figure.

relevant ethical regulations. Informed consent was obtained from all participants before beginning the study. Following the preregistration, we removed the 10 participants with the highest error rate from our analyses. Based on pilot studies and previous work (Cheyette & Piantadosi, 2020), we believed the sample size included for analysis (100 participants × 90 trials per participant = 9,000 data points) would be sufficient to determine model parameters within a small interval.

Materials

The experiment was designed in JavaScript using the psiTurk framework (Gureckis et al., 2016). There were 49 grid cells (7 × 7), with each grid cell 35 px^2 and an equal margin separating the cells. Unfilled grid cells were white and filled grid cells were gray with hex color No. A0A0A0. When a cell was clicked in the task, its border was bolded and turned red. The noise mask was multicolored static and had a size of 455 px^2 to cover the entire grid.

Design

There were four within-subject variables manipulated in the study: the number of cells filled (1–15); the exposure time of the displayed pattern (50 ms, 150 ms, 450 ms); and the direction of the changed cell from the first to the second presentation (white-to-gray or gray-to-white). Each multiple of number, time, and direction was shown exactly once, for a total of 15 × 3 × 2 = 90 trials. The initial direction of the changed cell was randomly chosen and then remained constant for the first 45 trials, with the last 45 trials assigned to the opposite direction. Within that constraint, the order of the trials was randomized, that is, number–time pairs were assigned randomly within each direction of change. The positions of the filled cells were chosen randomly on each trial. If the direction of change was white-to-gray, a random white cell from the initial exposure would turn gray on the second presentation; conversely, if the direction of change was gray-to-white, a random gray cell would turn white.

Procedure

After providing consent and reading instructions, participants began the first section of the experiment. Both halves of the experiment—the white-to-gray section and gray-to-white section—started with three practice trials. Participants were informed in both the practice trials and the main task whether a cell would be changing from white to gray, or vice versa. Each trial started with a fixation cross displayed on the center for 1,000 ms, followed by the grid with some cells filled in 50–450 ms, and then a noise mask for 600 ms. Then, the grid reappeared, with one modified cell. Subjects then clicked the cell they thought changed color and proceeded to the next trial. The basic setup is illustrated in Figure 2.

Results

We first ran a logistic regression predicting participants' accuracy from the number of gray cells, exposure time, and trial type ("appear" or "disappear"), which revealed significant effects of all three. There was a negative effect of the number shown ($B = -0.25$, $z = -38.6$, $p < .001$), such that more gray cells decreased accuracy, a positive effect of exposure duration ($B = 3.08$, $z = 20.3$, $p < .001$), and an effect of trial type such that participants performed better on trials where a cell appeared than disappeared ($B = 0.64$, $z = 12.5$, $p < .001$). The intercept was also significant ($B = 1.65$, $z = 24.3$, $p < .001$). Note that these analyses were performed post hoc (not preregistered) and run at the request of a reviewer.

We next fit the model to the data. To fit model parameters, we assumed that the information bound changes as a function of time according to a power law $B = a \cdot t^k$, where a and k are free parameters and t is exposure time in seconds. The other key parameter of the model is the weighting parameter in the loss function α, capturing the extent to which false negatives (high α) or false positives (low α) are more costly. To account for attention lapses and mis-presses, we also included a guessing-rate parameter, p_g, which captured the rate participants chose randomly from the set of valid alternatives (as opposed to via the model). We fit parameters under a hierarchical Bayesian model using Markov Chain Monte Carlo, assuming partial pooling of parameter estimates across participants (see Supplemental Materials).

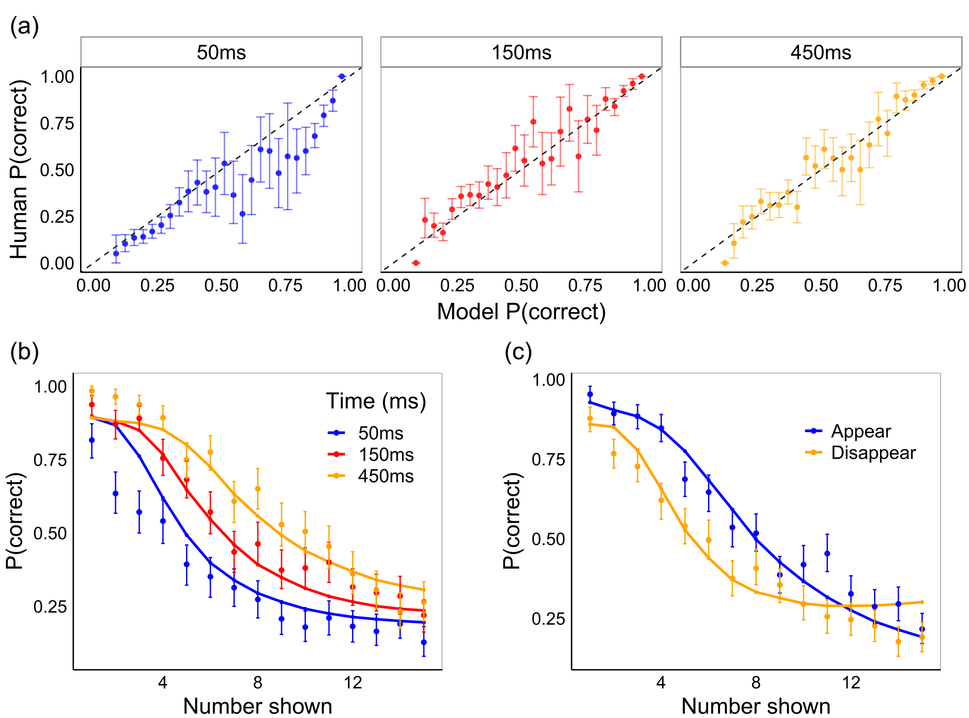
The maximum a posteriori (MAP) estimates for the group-level parameters were a = 33.5 (CI [32.2, 34.6]), k = 0.21 (CI [0.20, 0.22]), p_g = 0.16 (CI [0.12, 0.19]), and α = .35 (CI [0.33, 0.37]). This entails information bounds of 17.9, 22.5, and 28.3 bits at 50 ms, 150 ms, and 450 ms, respectively. The relatively high inferred rate of guessing likely reflects the fact that the model does not account for spatial errors, treating each cell independently. Figure 5a shows the model's predicted accuracy (*x*-axis) against human performance (*y*-axis) across all exposure durations (facets). Comparing the points to the dashed $y = x$ line reveals that the model's predictions tightly align with human accuracy across exposure durations, though the model is slightly biased to overestimate human performance at short times (left facet). The correlation between model predictions and human data across trials grouped by numerosity and exposure duration was 0.96 (R^2 = 0.93), another indication that the model provides a good fit to the data.

In line with participants, the model predicts near-veridical memory for visual displays with small numbers of objects, at longer exposure durations, and sharply increasing noise for larger numbers of objects and shorter durations. Figure 5b shows human accuracy (points and error bars), the model's predicted accuracy (lines) as a function of the total number of cells filled in, grouped by the exposure duration (colors). As predicted by the model, participants' performance saturates only for small numerosities at longer durations and quickly degrades as a function of number in each duration. The one notable discrepancy is that the model predicted better performance on small numerosities ($n < 4$) at 50 ms than was actually observed. Figure 5c depicts accuracy grouped by whether a cell appeared or disappeared from the first to second display and shows that participants performed substantially better on "appear" trials than "disappear" trials—a trend the model captures. The model would capture this trend even if α was fixed to 0.5, and in fact, higher values of α exaggerate rather than reduce the gap between "appear" and "disappear" trials.

To be clear, the fact that human performance on the change-localization task is strongly affected by numerosity is not an indication that the visual system is representing or using number. In fact, the experiment was explicitly designed so that number could not be used as a heuristic: Participants always knew that there was either one more gray cell on the second display than the first display (on "appear" trials) or one fewer (on "disappear" trials). Instead, the effect of numerosity on performance is an indication that spatial memory is making use of limited information in an efficient way,

Figure 5

Model Predictions and Data From Localization Task (Experiment 1)



Note. (a) Binned (25 bins/facet) model predictions (*x*-axis) and human data (*y*-axis) of performance on the change-localization task. Each facet shows predictions at different exposure durations. In (b) and (c) model, predictions are shown as lines, and human data from the change-localization task are shown as points with bootstrapped 95% confidence intervals. (b) Accuracy (*y*-axis) in the change-localization task as a function of the number of grid cells filled (*x*-axis) at each exposure duration. (c) Accuracy (*y*-axis) as a function of number (*x*-axis) grouped by whether or not a cell appeared or disappeared from first-to-second presentation. See the online article for the color version of this figure.

combining a prior expectation that there will be fewer gray pixels than white pixels with evidence gathered by observing the scene. Additionally, the inability to precisely remember scenes with more filled cells is a reflection of the fact that there are more ways to arrange scenes with more filled cells (up to half the number of grid cells), meaning that it takes more information to represent any one of them precisely.

Experiment 2

While Experiment 1 showed that the model is able to account for effects of number and exposure duration in spatial memory, it does not answer the question of whether human numerical estimation abilities arise from this same system. The goal of Experiment 2 is to replicate previously reported properties of number psychophysics and to test whether the model is able to capture these effects as well. If the patterns of noise (zero uncertainty then Weber's law) and bias (underestimation) in estimation derive from limitations in spatial memory, then the model of spatial memory should be able to explain the psychophysics of estimation at exposure durations (i.e., across different information capacities in visual memory); moreover, we should be able to recover similar parameter values from the model fit to a numerical estimation task as from the model fit to a spatial memory task. Alternatively, to the extent that the psychophysics of estimation derives from processing constraints that are independent of spatial memory, the visual memory model should not capture the psychophysics of estimation and the parameters recovered from model fitting should differ from those inferred in Experiment 1. To test these predictions, we ran a number estimation task with a design matched to Experiment 1.

Method

The procedure and display were identical to Experiment 1 up to the noise mask. After the noise mask, however, participants were asked to estimate the number of cells that were filled by typing a one- or two-digit numeric estimate into a text box. One hundred ten adult participants from Prolific again completed 90 trials, with each number (1–15) paired with duration (50 ms, 150 ms, 450 ms) displayed twice. Following the preregistration, we removed the 10 participants with the highest mean absolute error in estimation from our analyses and winsorized estimates to the 95% interval for each numerosity.

Results

We first ran linear regressions to predict participants' signed error (bias) and absolute error from the number of gray cells shown and the exposure duration. We found significant effects of both predictors in both cases. For signed errors, there was a positive intercept ($B = 0.89$, $t = 16.5$, $p < .001$), a negative effect of the number of gray cells ($B = -0.22$, $t = -43.69$, $p < .001$), and a positive effect of exposure duration ($B = 0.72$, $t = 5.76$, $p < .001$). This means that participants slightly overestimated small quantities at short exposures but increasingly underestimated larger quantities, and it means that the underestimation bias diminished with increasing exposure duration. For absolute errors, there was a negative intercept ($B = -0.17$, $t = -4.59$, $p < .001$), a positive effect of the number of gray cells ($B = 0.26$, $t = 76.64$, $p < .001$), and a negative effect of exposure duration ($B = -24.22$, $t = -24.22$, $p < .001$). This means participants became less accurate at estimating larger quantities and more accurate with longer exposures.

We fit the same parameters in the model with the estimation data as with the change-localization task. The MAP group-level parameters were $a = 32.9$ (CI [30.9, 33.8]), $k = 0.18$ (CI [0.17, 0.20]), $p_g = 0.03$ (CI [0.02, 0.03]), and $\alpha = .31$ (CI [0.29, 0.32]). The implied average information bounds are therefore 19.2, 23.4, and 28.5 bits at 50 ms, 150 ms, and 450 ms, respectively. This is slightly higher than the estimates derived from the change-localization task data, but the differences at each exposure duration are small (<10%). Table 1 provides a side-by-side comparison of the inferred MAP parameters from both experiments. A notable difference between the inferred parameters between the two tasks is the guessing rate, which is much lower than in the change-localization task. As noted previously, however, the relatively high guessing rate in the change-localization task is likely due to the fact that the model does not account for spatial errors or mis-presses (only completely random guessing)—this would increase the inferred rate of guessing in the change-localization task but not the estimation task.

The resulting psychophysical curves from the model (lines), along with the data from the experiment (points and error bars), are shown in Figure 6. The model captures the key psychophysical trends observed in the data: an underestimation bias that diminishes with exposure time; a subitizing range that increases with exposure time; scalar variability in estimation; and acuity in estimation that increases with exposure time. The nonzero but flat standard deviation for small numerosities in Figure 6b reflects the influence of guessing—without the guessing parameter it would show zero variability. The model predictions diverge somewhat from human performance on small numerosities ($n < 4$) at 50 ms—the model predicts better performance than is actually observed. An analogous discrepancy was observed in the change-localization task (also for $n < 4$ at 50 ms). It is possible that this occurs because the model assumes that participants each have a fixed guessing rate, when in fact people may be more likely to miss a display altogether at short exposure durations (e.g., if they are blinking).

Following the preregistration, we compared the model's maximum likelihood estimate (MLE) parameters for each subject to a standard psychophysical model of numerical estimation (Weber's law), as well as a modified one that accounts for the effects of time. The overall log likelihood of the model using MLE estimates of participants' parameters was −14,129. In the first comparison model, we assume that participants' estimates are drawn from a Gaussian centered around the number shown, n, with mean n and standard deviation $w \cdot n$, where w is a free parameter (their "Weber fraction"). We also fit a version of this where the standard deviation could vary as a function of time, such that $w = e^{w_0 + w_t \cdot t}$, where w_0 and w_t are fit and t is time in seconds. The median MLE w

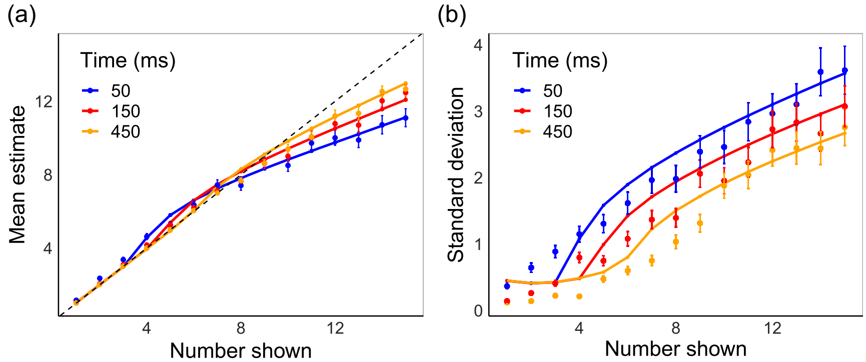
Table 1
MAP Parameters From Experiments 1 and 2

Experiment	a	k	α	p_g
Localization (E1)	33.5	0.21	.35	0.16
Estimation (E2)	32.9	0.18	.31	0.03

Note. MAP = maximum a posteriori; E = experiment.

Figure 6

Model Predictions and Data From Estimation Task (Experiment 2)



Note. (a) Mean estimates as a function of numerosity, grouped by exposure duration. (b) Standard deviations as a function of numerosity, grouped by exposure duration. Data is shown as points with 95% confidence intervals; the model fit to the human data shown as lines. See the online article for the color version of this figure.

fit in the static (non-time-varying) version was 0.24, with log likelihood −16,166. In the time-varying version, the median MLE w_0 was −1.15 and w_t was −1.75, giving w's of 0.29, 0.24, and 0.15 at 50 ms, 150 ms, and 450 ms, respectively, and had log likelihood −15,428. The Weber models thus did not fit nearly as well as our model, with Akaike information criterion differences of 3,974 and 2,498 (we preregistered Akaike information criterion differences of 10 as "significant").

Experiment 3

Experiments 1 and 2 demonstrate that a single model can fit human psychophysics of both a spatial localization task and a numerical estimation task. Furthermore, both the inferred parameters and the amount of information participants were inferred to have about the *spatial locations* of black dots and the *number* of black dots was numerically very close at each exposure duration, highly suggestive of a shared representational capacity and a common process. However, because Experiments 1 and 2 were between subjects, the most we can say is that the average parameter values recovered from both tasks are numerically close. To further assess whether there is a common process underlying both spatial localization and numerical estimation—and that our model provides a good account of that process—we ran a within-subjects experiment testing participants on both tasks. To the extent that participants' performance covaries across tasks and that this correlation is explained best by model parameters (other than participants' inferred guessing rate) in each task, this would provide stronger evidence of a common process.[2]

Method

We again collected data from 110 adult participants from Prolific, where each participant completed both the estimation task and the change-localization task. This was divided into three phases consisting of 45 trials each (135 total): Phase 1 was an estimation task identical in method to Experiment 2, except for half the total number of trials; Phase 2 was the change-localization task where a black cell always *appeared* between displays; and Phase 3 was the change-localization task where a black cell always *disappeared* (turned white) between displays. Phases 2 and 3 are identical in method to Experiment 1, except for that the order of appear and disappear trials was fixed. As in the previous experiments, we removed the 10 lowest performing participants, which this time was determined as the five participants with the highest average absolute error on the estimation task and the five participants with the highest average error on the localization task (after removing the first five).

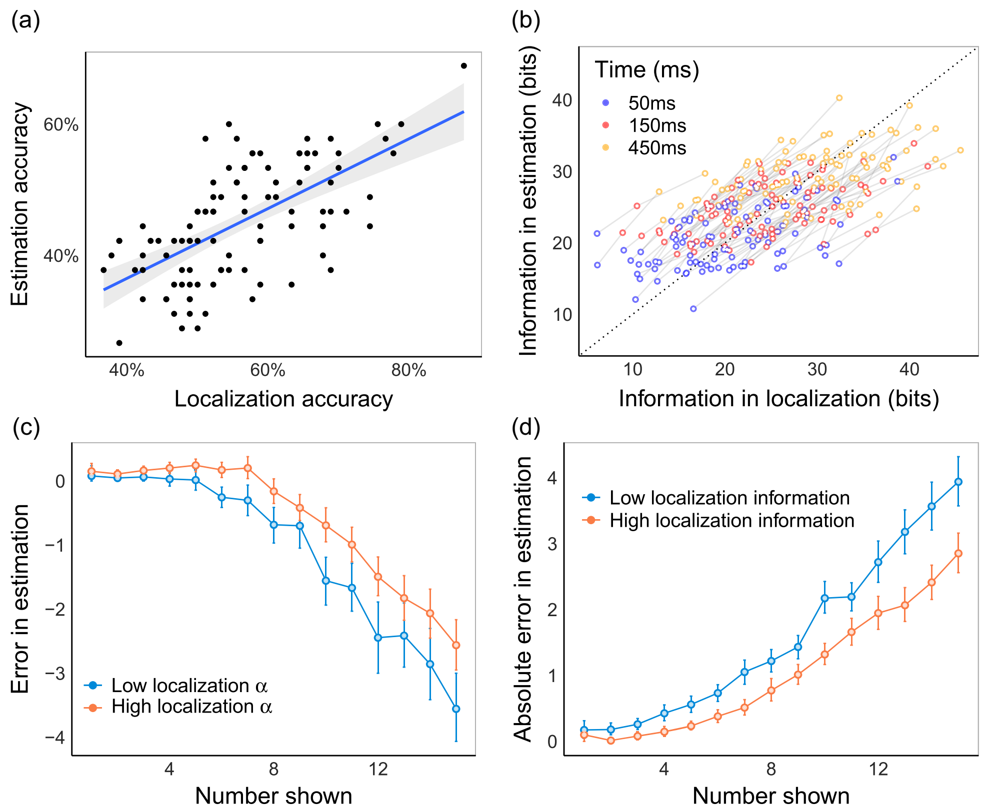
Results

We first ran a regression predicting participants' mean accuracy across all trials in the estimation task from their mean accuracy in the localization task, which revealed a strong correlation, $r = 0.63$ ($p <$ .001). This is shown in Figure 7a. This is notably high, given that the correlation between participants' mean accuracy on *appear* and *disappear* trials of the change-localization portion of the experiment was numerically nearly identical, $r = 0.61$ ($p < .001$). One obvious concern is that these high correlations may be mostly or entirely driven by differences in attention or motivation. We therefore reran a regression on only the participants who had above the 50th percentile overall accuracy in estimation. This again revealed a correlation between localization performance and estimation performance, $r = 0.51$ ($p < .001$), which was again numerically similar to the correlation between performance on *appear* and *disappear* trials of the localization task, $r = 0.56$ ($p < .001$). Differences in attention therefore seem unlikely to explain the observed relationship between estimation and spatial localization performance.

We next found MLE parameter estimates under the model given each participant's estimation and change-localization data (separating data from the two tasks). The mean parameter estimates for both tasks are shown in Table 2. We tested the extent to which participants' inferred parameters were consistent across tasks and found significant positive correlations for all inferred parameters

[2] This experiment and the analyses performed were run at the suggestion of a reviewer and not part of the preregistration.

Figure 7
Data From Experiment 3



Note. (a) Participants' average accuracy in the localization task averaged across all trials (*x*-axis) versus their average accuracy in the estimation task (*y*-axis). (b) Each participant's inferred information in bits at each exposure time (color), fit to their localization task data (*x*-axis) and their estimation task data (*y*-axis). (c) Bias in numeric estimation, grouped by participants who were inferred to have high (orange) and low (blue) α in the localization task, using a median split. (d) Absolute estimation error, grouped by high (orange) and low (blue) inferred information capacity in localization. Error bars are 95% confidence interval. See the online article for the color version of this figure.

apart from guessing rate[3]: $r = 0.45$ ($p < .001$) for a; $r = 0.49$ ($p < .001$) for k; $r = 0.44$ ($p < .001$) for α; and $r = -0.05$ ($p < .062$) for p_g. This means that, for example, we can predict the degree to which someone will underestimate their loss parameter α inferred in the localization task (Figure 7c). Participants' inferred information capacities over time were also highly correlated across tasks (Figure 7b): $r = 0.65$ ($p < .001$). This was also true of inferred information capacities *within* each exposure duration: $r = 0.54$ at 50 ms; $r = 0.52$ at 150 ms; and $r = 0.48$ at 450 ms ($p < .001$). Figure 7c and 7d illustrates two ways that inferred latent parameters from the localization task predict behavioral differences in estimation. Figure 7c shows how bias in participants' numeric estimation (i.e., signed error from the true numerosity) is predicted by the α value inferred from their localization task performance, such that those with lower inferred α (<50th percentile) underestimate more than those with higher inferred α (≥50th percentile). Figure 7d shows how absolute estimation error (i.e., absolute deviation from the true numerosity) is modulated by the information capacity inferred from the localization task, such that lower (<50th percentile) inferred information capacity bounds result in higher absolute error than higher (≥50th percentile) inferred capacity bounds.

We also tested whether the estimated parameter values for individual participants differed across tasks by running paired *t* tests on their differences. The inferred parameter values from the estimation task did not differ significantly from the inferred parameter values from the localization task for either a, $t(99) = -0.06$, $p = .95$, or k, $t(99) = -1.25$, $p = .21$, which control how people accumulate information over time. The loss function parameter α was inferred to be lower in the estimation task (μ = 0.26) than in the localization task

Table 2
Mean MLE Parameters From Experiment 3

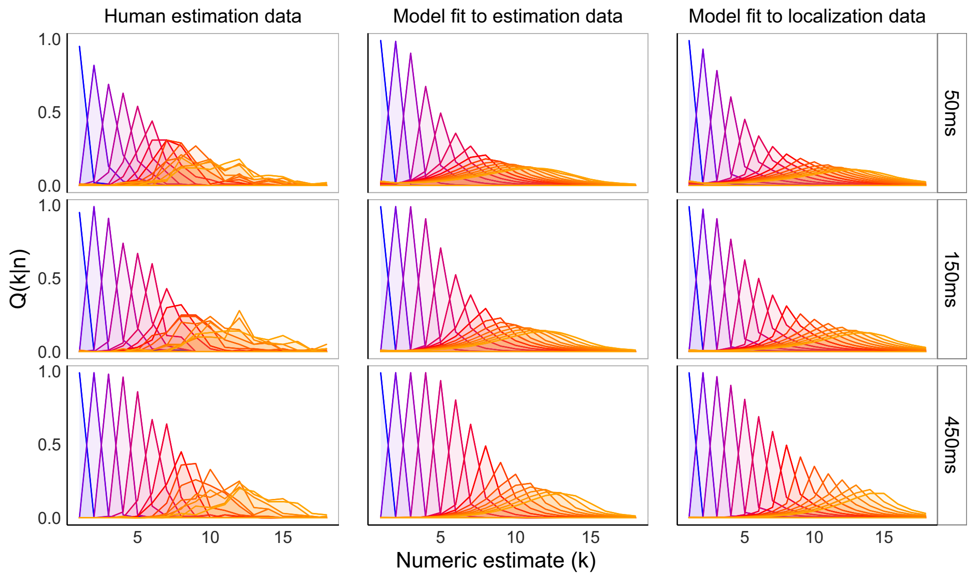
Task	a	k	α	p_g
Localization	33.5	0.17	.33	0.15
Estimation	33.4	0.16	.26	0.03

Note. MLE = maximum likelihood estimate.

[3] We reran all these analyses without removing the 10 worst-performing participants, which gives nearly identical results (and no qualitative differences).

Figure 8

Group-Level Psychophysical Curves From Estimation and Localization



Note. Group-level psychophysical curves showing the proportion (*y*-axis) of estimates *k* (*x*-axis) for each numerosity presented $n = 1 \ldots 15$ (lines) at each exposure time tested in Experiment 3 (rows). The leftmost column is the true human estimation data; the middle column shows the model fit to the data from the estimation task; and the rightmost column shows the model fit to the data from the localization task. See the online article for the color version of this figure.

($\mu = 0.33$), a difference that was significant, $t(99) = -8.38$, $p < .001$, as was the inferred guessing rate, $t(99) = -10.9$, $p < .001$. Despite the small numeric difference in inferred α, the fit to the localization task predicts number psychophysics that are remarkably consistent at both the individual and group levels. Figure 8 highlights the alignment at the group level, showing aggregate response distributions from the estimation task on the left, average response distributions from the model fit to participants' estimation task data in the middle, and average response distributions fit to participants' localization task data on the right.

Discussion

This article presented a model of visuospatial memory that captures human performance both in a spatial memory task and in a quantity estimation task. Crucially, many key properties of numerical cognition—including a transition from exactness to approximation, roughly Gaussian response distributions, underestimation, and effects of time—can be recovered from a nonnumerical visual task using a model optimized to remember items' locations. Furthermore, we were able to predict the degree of bias (under- or overestimation) and noise in a person's numerical estimates by fitting model parameters to their performance in the spatial localization task. Our results therefore indicate that the psychophysics of number in vision can largely be attributed to uncertainty regarding the items displayed in a scene, rather than to number-specific processing. While there must exist some number-specific processing—quantity must be extracted from visual memory—our findings indicate that Weber's law, subitizing, underestimation, and other effects observed in numerical estimation are not the direct *result* of that processing.

Because the model accounts for subitizing as well as large number estimation, it also demonstrates how a single mechanism might give rise to the observed behavioral discontinuities between large and small numbers. This is because the model predicts different patterns of behavior above and below its capacity limit—visual representations are exact and perfect only when scenes are simple. After that, a bounded-optimal perceptual system Q exhibits the known properties of large number estimation, which arise here from imperfectly individuating objects and tracking their locations. Finally, although the large number system is commonly thought to represent analog magnitudes on a continuous scale (Carey, 2009; Feigenson et al., 2004), the model demonstrates how noisy beliefs over discrete representations can give rise to what appears to be analog behavior (see also Beck, 2015, Clarke & Beck, 2021, and Clarke, 2022, for philosophical treatments related to this point).

Some studies have found a strong relationship between object-tracking ability, visual memory capacity, and estimation acuity outside the subitizing range, as predicted by our model (Bugden & Ansari, 2016; Green & Bavelier, 2003, 2006; Passolunghi et al., 2015). However, other studies have found a stronger link between an individual's visual working memory capacity and their subitizing range than with their estimation acuity (Piazza et al., 2011; Revkin et al., 2008), which seems to contradict predictions of our theory or conflict with the results of Experiment 3. Importantly, though, while the model does link both subitizing range and estimation to visuospatial information capacity, differences in information capacity do not necessarily cause equally large changes to the subitizing range and estimation acuity. Specifically, modulating the information bound tends to affect the subitizing range substantially more than the (implicit) Weber fraction (see Supplemental Materials). Furthermore, the tasks meant to index visual working memory capacity employed in those studies are subtly different than the one used here. Whereas Revkin et al. (2008) and Piazza et al. (2011) used a change-detection paradigm in which a property of one of the objects in the display (e.g., color) might change, our task involves tracking only the presence or absence of objects at particular locations—not properties (like color) bound to the objects. The task we employed, therefore, may better index the specifically *spatial* component of visual memory we believe underlies both individuation and subsequent enumeration (Pylyshyn, 1989; Trick & Pylyshyn, 1993, 1994).

Another challenge for our proposal is that it does not seem consistent with the apparent failure observed in some cases when infants and young children are asked to compare small quantities (1–3) against large quantities (4+), Feigenson and Carey (2003), Feigenson et al. (2002), Lipton and Spelke (2004), and Xu (2003); though see Cordes and Brannon (2009a, 2009b), Mack (2006), and Strauss and Curtis (1981) for conflicting evidence. For instance, Feigenson et al. (2002) ran a manual search task where two containers were baited with crackers and found that infants crawled toward the container with more crackers when there were one versus two crackers and two versus three crackers but not two versus four crackers or three versus six crackers. These data have been taken as evidence of two separate systems for processing small and large quantities (Feigenson et al., 2004). However, there are two reasons these findings do not necessarily conflict with our account. First, given a low information capacity bound, the model predicts higher performance comparing one versus two objects relative to two versus four objects and three versus six objects; and although the model would predict success in this case at, for example, comparisons of one versus four and two versus eight, there is also evidence that young infants discriminate these higher ratio quantities (Cordes & Brannon, 2009a, 2009b). In fact, we fit the model to the data presented in Feigenson et al. (2002) and found that its predictions are compatible (see Supplemental Figure S7), falling well within the 95% CI for each comparison tested. Second, our claim is not that there is one rather than two systems. Instead, we are proposing that the system for individuating objects is the input to the system for computing numerosity. So, a simple extension of the present model would be to suppose that, in infants, the individuation system does not pass on information to the system for representing quantities in cases of zero uncertainty.

In a previous article, we found that people underestimate less and become more precise in estimation as they make saccades across a scene containing a large (10–80) number of items (Cheyette & Piantadosi, 2019). We interpreted this as people accumulating an approximate count of objects in their visual path and not counting a significant proportion of peripherally viewed objects. While our account in this article was not mechanistic—and was not intended to be—our findings here present an alternative interpretation of the earlier results: that people were actually accumulating *spatial* information about where items were located in the display, and this resulted in downstream improvements in numerosity judgments. An important future direction is therefore linking functional-level accounts, like the one presented in this article, to mechanistic models of the visual routines involved in object tracking and estimation. The mechanics of visual attention may be necessary to explain the effects of item arrangement and grouping (e.g., Anobile et al., 2020; Atkinson et al., 1976; Ciccione & Dehaene, 2020; Ginsburg, 1976; Krajcsi et al., 2013; G. S. Starkey & McCandliss, 2014; Trick & Enns, 1997; Van Oeffelen & Vos, 1982), though some effects relating to complexity and regularity—such as an increased subitizing range from canonical displays (e.g., Mandler & Shebo, 1982)—might be explained in terms of "ease of encoding" using information-theoretic methods like the ones employed in this article. Similarly, although our account is broadly consistent with studies showing that numerosity judgments depend on object segmentation (Franconeri et al., 2009), our model provides no way of describing how or when a group may be perceived as a single entity and how that subsequently affects representations of quantity.

It is worth highlighting two other important limitations of our model and experiments that leave room for future work. First, the model and experiments were only designed to capture numerical perception in the domain of vision. However, innate numerical abilities have been documented in audition, touch, and across modalities (Barth et al., 2003; Mix et al., 1997; Plaisier et al., 2009; P. Starkey et al., 1990). Though the model we presented here was designed to deal with spatial rather than temporal integration (Meck & Church, 1983), we believe similar principles of information processing are likely to apply and hence the methods used in this article could be adapted to capture, for example, the processing of auditory sequences (Cheatham & White, 1954; Izard et al., 2009). The other main limitation is our use of simplifying assumptions to model spatial memory—specifically, in discretizing the space and in assuming objects to be identical. The model would thus need to be extended to capture, for instance, the influences of continuous visual features such as surface area, convex hull, and density on numerosity perception (e.g., Cantrell et al., 2015; Cantrell & Smith, 2013; Gebuis et al., 2016; Gebuis & Reynvoet, 2012; Leibovich et al., 2017; Lourenco & Longo, 2010, 2011; Mix et al., 2002; Newcombe et al., 2015; Sokolowski et al., 2017). In fact, the methods we employed in this article may be useful to understand some of these effects: Because continuous features like surface area are correlated with numerosity in the real world, principles of efficient information compression dictate that their representations will not be independent.

References

Alonso-Diaz, S., Cantlon, J. F., & Piantadosi, S. T. (2018). A threshold-free model of numerosity comparisons. *PLOS ONE*, *13*(4), Article e0195188. https://doi.org/10.1371/journal.pone.0195188

Alvarez, G. A., & Franconeri, S. L. (2007). How many objects can you track?: Evidence for a resource-limited attentive tracking mechanism. *Journal of Vision*, *7*(13), Article 14. https://doi.org/10.1167/7.13.14

Anobile, G., Castaldi, E., Moscoso, P. A. M., Burr, D. C., & Arrighi, R. (2020). "Groupitizing": A strategy for numerosity estimation. *Scientific Reports*, *10*(1), Article 13436. https://doi.org/10.1038/s41598-020-68111-1

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

Barth, H., Kanwisher, N., & Spelke, E. (2003). The construction of large number representations in adults. *Cognition*, *86*(3), 201–221. https://doi.org/10.1016/s0010-0277(02)00178-6

Beck, J. (2015). Analogue magnitude representations: A philosophical introduction. *The British Journal for the Philosophy of Science*, *66*(4), 829–855. https://doi.org/10.1093/bjps/axu014

Bugden, S., & Ansari, D. (2016). Probing the nature of deficits in the 'approximate number system'in children with persistent developmental dyscalculia. *Developmental Science*, *19*(5), 817–833. https://doi.org/10.1111/desc.12324

Cantrell, L., Boyer, T. W., Cordes, S., & Smith, L. B. (2015). Signal clarity: An account of the variability in infant quantity discrimination tasks. *Developmental Science*, *18*(6), 877–893. https://doi.org/10.1111/desc.12283

Cantrell, L., & Smith, L. B. (2013). Open questions and a proposal: A critical review of the evidence on infant numerical abilities. *Cognition*, *128*(3), 331–352. https://doi.org/10.1016/j.cognition.2013.04.008

Carey, S. (2009). *The origin of concepts*. Oxford University Press.

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

Cheyette, S. J., & Piantadosi, S. T. (2019). A primarily serial, foveal accumulator underlies approximate numerical estimation. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(36), 17729–17734. https://doi.org/10.1073/pnas.1819956116

Cheyette, S. J., & Piantadosi, S. T. (2020). A unified account of numerosity perception. *Nature Human Behaviour*, *4*(12), 1265–1272. https://doi.org/10.1038/s41562-020-00946-0

Ciccione, L., & Dehaene, S. (2020). Grouping mechanisms in numerosity perception. *Open Mind*, *4*, 102–118. https://doi.org/10.1162/opmi_a_00037

Clarke, S. (2022). Beyond the icon: Core cognition and the bounds of perception. *Mind & Language*, *37*(1), 94–113. https://doi.org/10.1111/mila.12315

Clarke, S., & Beck, J. (2021). The number sense represents (rational) numbers. *Behavioral and Brain Sciences*, *44*, Article e178. https://doi.org/10.1017/S0140525X21000571

Cordes, S., & Brannon, E. M. (2009a). Crossing the divide: Infants discriminate small from large numerosities. *Developmental Psychology*, *45*(6), Article 1583. https://doi.org/10.1037/a0015666

Cordes, S., & Brannon, E. M. (2009b). The relative salience of discrete and continuous quantity in young infants. *Developmental Science*, *12*(3), 453–463. https://doi.org/10.1111/j.1467-7687.2008.00781.x

Dehaene, S. (2011). *The number sense: How the mind creates mathematics*. Oxford University Press.

Dehaene, S., & Changeux, J.-P. (1993). Development of elementary numerical abilities: A neuronal model. *Journal of Cognitive Neuroscience*, *5*(4), 390–407. https://doi.org/10.1162/jocn.1993.5.4.390

Dehaene, S., & Mehler, J. (1992). Cross-linguistic regularities in the frequency of number words. *Cognition*, *43*(1), 1–29. https://doi.org/10.1016/0010-0277(92)90030-l

DeWind, N. K., Bonner, M. F., & Brannon, E. M. (2020). Similarly oriented objects appear more numerous. *Journal of Vision*, *20*(4), Article 4. https://doi.org/10.1167/jov.20.4.4

Feigenson, L., & Carey, S. (2003). Tracking individuals via object-files: Evidence from infants' manual search. *Developmental Science*, *6*(5), 568–584. https://doi.org/10.1111/1467-7687.00313

Feigenson, L., Carey, S., & Hauser, M. (2002). The representations underlying infants' choice of more: Object files versus analog magnitudes. *Psychological Science*, *13*(2), 150–156. https://doi.org/10.1111/1467-9280.00427

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences*, *8*(7), 307–314. https://doi.org/10.1016/j.tics.2004.05.002

Franconeri, S. L., Bemis, D. K., & Alvarez, G. A. (2009). Number estimation relies on a set of segmented objects. *Cognition*, *113*(1), 1–13. https://doi.org/10.1016/j.cognition.2009.07.002

Gebuis, T., Kadosh, R. C., & Gevers, W. (2016). Sensory-integration system rather than approximate number system underlies numerosity processing: A critical review. *Acta Psychologica*, *171*, 17–35. https://doi.org/10.1016/j.actpsy.2016.09.003

Gebuis, T., & Reynvoet, B. (2012). The interplay between nonsymbolic number and its continuous visual properties. *Journal of Experimental Psychology: General*, *141*(4), 642–648. https://doi.org/10.1037/a0026218

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*(4), 457–472. https://doi.org/10.1214/ss/1177011136

Ginsburg, N. (1976). Effect of item arrangement on perceived numerosity: Randomness vs regularity. *Perceptual and Motor Skills*, *43*(2), 663–668. https://doi.org/10.2466/pms.1976.43.2.663

Green, C. S., & Bavelier, D. (2003). Action video game modifies visual selective attention. *Nature*, *423*(6939), 534–537. https://doi.org/10.1038/nature01647

Green, C. S., & Bavelier, D. (2006). Enumeration versus multiple object tracking: The case of action video game players. *Cognition*, *101*(1), 217–245. https://doi.org/10.1016/j.cognition.2005.10.004

Gureckis, T. M., Martin, J., McDonnell, J., Rich, A. S., Markant, D., Coenen, A., Halpern, D., Hamrick, J. B., & Chan, P. (2016). Psiturk: An open-source framework for conducting replicable behavioral experiments online. *Behavior Research Methods*, *48*(3), 829–842. https://doi.org/10.3758/s13428-015-0642-8

Haladjian, H. H., & Pylyshyn, Z. W. (2011). Enumerating by pointing to locations: A new method for measuring the numerosity of visual object representations. *Attention, Perception, & Psychophysics*, *73*, 303–308. https://doi.org/10.3758/s13414-010-0030-5

Heng, J. A., Woodford, M., & Polania, R. (2020). Efficient sampling and noisy decisions. *eLife*, *9*, Article e54962. https://doi.org/10.7554/eLife.54962

Inglis, M., & Gilmore, C. (2013). Sampling from the mental number line: How are approximate number system representations formed? *Cognition*, *129*(1), 63–69. https://doi.org/10.1016/j.cognition.2013.06.003

Izard, V., & Dehaene, S. (2008). Calibrating the mental number line. *Cognition*, *106*(3), 1221–1247. https://doi.org/10.1016/j.cognition.2007.06.004

Izard, V., Sann, C., Spelke, E. S., & Streri, A. (2009). Newborn infants perceive abstract numbers. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(25), 10382–10385. https://doi.org/10.1073/pnas.0812142106

Jevons, W. S. (1871). The power of numerical discrimination. *Nature*, *3*, 281–282. https://doi.org/10.1038/003281a0

Kaufman, E. L., Lord, M. W., Reese, T. W., & Volkmann, J. (1949). The discrimination of visual number. *The American Journal of Psychology*, *62*(4), 498–525. https://doi.org/10.2307/1418556

Kim, G., Jang, J., Baek, S., Song, M., & Paik, S.-B. (2021). Visual number sense in untrained deep neural networks. *Science Advances*, *7*(1), Article eabd6127. https://doi.org/10.1126/sciadv.abd6127

Krajcsi, A., Szabó, E., & Mórocz, I. Á. (2013). Subitizing is sensitive to the arrangement of objects. *Experimental Psychology*, *60*(4), 227–237. https://doi.org/10.1027/1618-3169/a000191

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*, Article e164. https://doi.org/10.1017/S0140525X16000960

Lipton, J. S., & Spelke, E. S. (2004). Discrimination of large and small numerosities by human infants. *Infancy*, *5*(3), 271–290. https://doi.org/10.1111/desc.12108

Lourenco, S. F., & Longo, M. R. (2010). General magnitude representation in human infants. *Psychological Science*, *21*(6), 873–881. https://doi.org/10.1177/0956797610370158

Lourenco, S. F., & Longo, M. R. (2011). *Origins and development of generalized magnitude representation*. Elsevier.

Ma, W. J., Husain, M., & Bays, P. M. (2014). Changing concepts of working memory. *Nature Neuroscience*, *17*(3), 347–356. https://doi.org/10.1038/nn.3655

Mack, W. (2006). Numerosity discrimination: Infants discriminate small from large numerosities. *European Journal of Developmental Psychology*, *3*(1), 31–47. https://doi.org/10.1080/17405620500347695

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

Meck, W. H., & Church, R. M. (1983). A mode control model of counting and timing processes. *Journal of Experimental Psychology: Animal Behavior Processes*, *9*(3), Article 320. https://doi.org/10.1037/0097-7403.9.3.320

Mix, K. S., Huttenlocher, J., & Levine, S. C. (2002). Multiple cues for quantification in infancy: Is number one of them? *Psychological Bulletin*, *128*(2), 278–294. https://doi.org/10.1037/0033-2909.128.2.278

Mix, K. S., Levine, S. C., & Huttenlocher, J. (1997). Numerical abstraction in infants: Another look. *Developmental Psychology*, *33*(3), 423–428. https://doi.org/10.1037//0012-1649.33.3.423

Newcombe, N. S., Levine, S. C., & Mix, K. S. (2015). Thinking about quantity: The intertwined development of spatial and numerical cognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, *6*(6), 491–505. https://doi.org/10.1002/wcs.1369

Nieder, A., & Dehaene, S. (2009). Representation of number in the brain. *Annual Review of Neuroscience*, *32*, 185–208. https://doi.org/10.1146/annurev.neuro.051508.135550

Nieder, A., & Miller, E. K. (2004). Analog numerical representations in rhesus monkeys: Evidence for parallel processing. *Journal of Cognitive Neuroscience*, *16*(5), 889–901. https://doi.org/10.1162/089892904970807

Palomares, M., Smith, P. R., Pitts, C. H., & Carter, B. M. (2011). The effect of viewing eccentricity on enumeration. *PLOS ONE*, *6*(6), Article e20779. https://doi.org/10.1371/journal.pone.0020779

Passolunghi, M. C., Lanfranchi, S., Altoè, G., & Sollazzo, N. (2015). Early numerical abilities and cognitive skills in kindergarten children. *Journal of Experimental Child Psychology*, *135*, 25–42. https://doi.org/10.1016/j.jecp.2015.02.001

Piantadosi, S. T., & Cantlon, J. F. (2017). True numerical cognition in the wild. *Psychological Science*, *28*(4), 462–469. https://doi.org/10.1177/0956797616686862

Piazza, M., Fumarola, A., Chinello, A., & Melcher, D. (2011). Subitizing reflects visuo-spatial object individuation capacity. *Cognition*, *121*(1), 147–153. https://doi.org/10.1016/j.cognition.2011.05.007

Pica, P., Lemer, C., Izard, V., & Dehaene, S. (2004). Exact and approximate arithmetic in an amazonian indigene group. *Science*, *306*(5695), 499–503. https://doi.org/10.1126/science.1102085

Plaisier, M. A., Tiest, W. M. B., & Kappers, A. M. (2009). One, two, three, many–subitizing in active touch. *Acta Psychologica*, *131*(2), 163–170. https://doi.org/10.1016/j.actpsy.2009.04.003

Pylyshyn, Z. (1989). The role of location indexes in spatial perception: A sketch of the FINST spatial-index model. *Cognition*, *32*(1), 65–97. https://doi.org/10.1016/0010-0277(89)90014-0

Qu, C., DeWind, N. K., & Brannon, E. M. (2022). Increasing entropy reduces perceived numerosity throughout the lifespan. *Cognition*, *225*, Article 105096. https://doi.org/10.1016/j.cognition.2022.105096

Revkin, S. K., Piazza, M., Izard, V., Cohen, L., & Dehaene, S. (2008). Does subitizing reflect numerical estimation? *Psychological Science*, *19*(6), 607–614. https://doi.org/10.1111/j.1467-9280.2008.02130.x

Sokolowski, H. M., Fias, W., Ononye, C. B., & Ansari, D. (2017). Are numbers grounded in a general magnitude processing system? a functional neuroimaging meta-analysis. *Neuropsychologia*, *105*, 50–69. https://doi.org/10.1016/j.neuropsychologia.2017.01.019

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

Starkey, P., Spelke, E. S., & Gelman, R. (1990). Numerical abstraction by human infants. *Cognition*, *36*(2), 97–127. https://doi.org/10.1016/0010-0277(90)90001-z

Stoianov, I., & Zorzi, M. (2012). Emergence of a'visual number sense'in hierarchical generative models. *Nature Neuroscience*, *15*(2), 194–196. https://doi.org/10.1038/nn.2996

Strauss, M. S., & Curtis, L. E. (1981). Infant perception of numerosity. *Child Development*, 1146–1152. https://doi.org/10.2307/1129500

Testolin, A., Dolfi, S., Rochus, M., & Zorzi, M. (2020). Visual sense of number vs. sense of magnitude in humans and machines. *Scientific Reports*, *10*(1), Article 10045. https://doi.org/10.1038/s41598-020-66838-5

Testolin, A., Zou, W. Y., & McClelland, J. L. (2020). Numerosity discrimination in deep neural networks: Initial competence, developmental refinement and experience statistics. *Developmental Science*, *23*(5), Article e12940. https://doi.org/10.1111/desc.12940

Trick, L. M., & Enns, J. T. (1997). Clusters precede shapes in perceptual organization. *Psychological Science*, *8*(2), 124–129. https://doi.org/10.1111/j.1467-9280.1997.tb00694.x

Trick, L. M., & Pylyshyn, Z. W. (1993). What enumeration studies can show us about spatial attention: Evidence for limited capacity preattentive processing. *Journal of Experimental Psychology: Human Perception and Performance*, *19*(2), 331–351. https://doi.org/10.1037//0096-1523.19.2.331

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*(1), 80–102. https://doi.org/10.1037/0033-295x.101.1.80

Van den Berg, R., Shin, H., Chou, W.-C., George, R., & Ma, W. J. (2012). Variability in encoding precision accounts for visual short-term memory limitations. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(22), 8780–8785. https://doi.org/10.1073/pnas.1117465109

Van Oeffelen, M. P., & Vos, P. G. (1982). Configurational effects on the enumeration of dots: Counting by groups. *Memory & Cognition*, *10*, 396–404. https://doi.org/10.3758/bf03202432

Vul, E., Frank, M. C., Tenenbaum, J. B., & Alvarez, G. (2009). *Explaining human multiple object tracking as resource-constrained approximate inference in a dynamic probabilistic model* [Conference session]. Proceedings of the 22nd international conference on neural information processing systems (pp. 1955–1963).

Woodford, M. (2020). Modeling imprecision in perception, valuation, and choice. *Annual Review of Economics*, *12*, 579–601. https://doi.org/10.1146/annurev-economics-102819-040518

Xu, F. (2003). Numerosity discrimination in infants: Evidence for two systems of representations. *Cognition*, *89*(1), B15–B25. https://doi.org/10.1016/s0010-0277(03)00050-7

Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition*, *74*(1), B1–B11. https://doi.org/10.1016/s0010-0277(99)00066-9

Yang, T.-I., & Chiao, C.-C. (2016). Number sense and state-dependent valuation in cuttlefish. *Proceedings of the Royal Society B: Biological Sciences*, *283*(1837), Article 20161379. https://doi.org/10.1098/rspb.2016.1379

Zorzi, M., & Testolin, A. (2018). An emergentist perspective on the origin of number sense. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1740), Article 20170043. https://doi.org/10.1098/rstb.2017.0043

Received October 11, 2022
Revision received January 8, 2024
Accepted January 24, 2024 ■