The psychophysics of number arise from resource-limited spatial memory

Samuel J. Cheyette (sam_cheyette@berkeley.edu)
Department of Psychology, UC Berkeley

Shengyi Wu (shengyiwu@berkeley.edu)
Department of Psychology, UC Berkeley

Steven T. Piantadosi (stp@berkeley.edu)
Department of Psychology, UC Berkeley

Abstract

People can identify the number of objects in small sets rapidly and without error but become increasingly noisy for larger sets. However, the cognitive mechanisms underlying these ubiquitous psychophysics are poorly understood. We present a model of a limited-capacity visual system optimized to individuate and remember the location of objects in a scene which gives rise to all key aspects of number psychophysics, including error-free small number perception and scalar variability for larger numbers. We therefore propose that number psychophysics can be understood as an emergent property of primitive perceptual mechanisms — namely, the process of identifying and representing individual objects in a scene. To test our theory, we ran two experiments: a change-localization task to measure participants' memory for the locations of objects (Experiment 1) and a numerical estimation task (Experiment 2). Our model accounts well for participants' performance in both experiments, despite only being optimized to efficiently encode where objects are present in a scene. Our results demonstrate that the key psychophysical features of numerical cognition do not arise from separate modules or capacities specific to number, but rather from lower-level constraints on perception which are manifested even in non-numerical tasks.

Introduction

Numerosity perception has been studied for at least 150 years (Jevons, 1871) and its psychophysics have been well characterized. Most notably, for small sets of up to about four objects, people are error-less; above that, error scales roughly linearly with numerosity (Feigenson et al., 2004; Jevons, 1871; Revkin et al., 2008). Weber fractions, which describe the rate at which internal noise scales with numerosity, are by far the most common measure in the field of numerical cognition. However, while providing a good descriptive characterization of number psychophysics, Weber fractions are entirely agnostic to mechanism — i.e., it is entirely unclear what cognitive limitations are responsible for the internal noise. While there has been interest in characterizing what visual features people may rely on to estimate numerosity (e.g. Anobile et al., 2018; Gebuis et al., 2016; Lourenco, 2015; Lourenco & Longo, 2011; Sokolowski et al., 2017), there have been only a handful of formal models aimed at understanding how people convert a scene into the summary statistic of number (Dehaene & Changeux, 1993; Stoianov & Zorzi, 2012; Testolin, Dolfi, et al., 2020; Testolin, Zou, et al., 2020).

These recent models of numerosity perception have largely used unsupervised neural networks that learn to extract statistical features of visual scenes; linear classifiers are then trained on hidden layer representations to test latent numerical discrimination ability (Stoianov & Zorzi, 2012; Testolin, Dolfi, et al., 2020; Testolin, Zou, et al., 2020). These models have demonstrated that numerosity is naturally extracted as a useful statistical component of visual scenes; and, like humans, the networks have representations of number that are intertwined with other correlated visual dimensions such as the surface area or density of objects. They also reproduce the signature scalar variability observed in large number estimation, though none of them produces the phenomenon of error-free "subitizing".

While it is widely believed that subitizing and large number estimation involve separate cognitive processes (Carey, 2009; Feigenson et al., 2004), other recent modeling work of ours has called this theory into question. In a recent paper, we demonstrated that any optimal but resource-constrained system should demonstrate a discontinuity in estimation ability from zero error for small numbers to scalar variability beyond (Cheyette & Piantadosi, 2020). Furthermore, we showed that the model predicts many aspects of human number psychophysics — including when subitizing should transition to estimation — at different amounts of available visual information. However, this model only explains in theory how a bounded-optimal number system *should* behave — but neither this nor any other model has demonstrated what cognitive mechanisms actually produce number psychophysics.

In this paper, we propose that number psychophysics arise from the process of identifying and representing individual objects, a theory we formalize in a computational model of bandwidth-limited scene memory. Given a scene as input, the model 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. Even though the model is explicitly optimized only to detect and remember the presence of objects in various locations, we show that the resultant probability distributions over numerosities nonetheless predict many key properties of number psychophysics, including both subitizing and scalar variability. The full set of model predictions that match previously reported findings in human numerosity perception include: i) exact or near-exact estimation of small sets (subitizing) (Jevons, 1871; Kaufman et al., 1949; Mandler & Shebo, 1982); ii) a subitizing range that increases as a function of display time (Cheyette & Piantadosi, 2020); iii) roughly normally-shaped response distributions for estimation (Nieder & Dehaene, 2009; Pica et al., 2004); iv) increasingly noisy estimation (scalar variability) for larger sets (Dehaene, 2011; Feigenson et al., 2004; Jevons, 1871); v) estima-

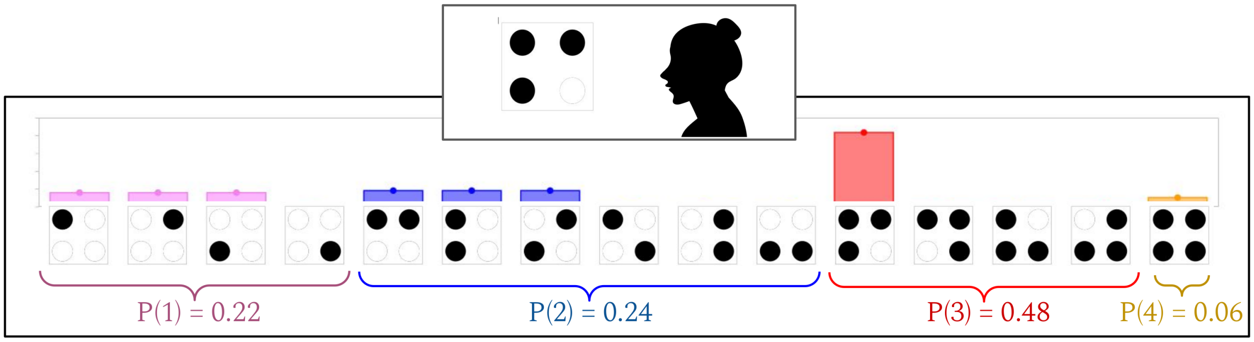


Figure 1: This figure conceptually illustrates how the model works, simplifying it to assume that there are only 4 pixels for clarity. In this example, a person sees a scene with 3 objects, which is represented as a probability distribution over all possibilities of what she saw. Possible arrangements 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.

tion acuity that increases with display time (Inglis & Gilmore, 2013); and vi) an under-estimation bias for large sets that diminishes with increased display time (Cheyette & Piantadosi, 2019, 2020).

At higher information capacity bounds, the model predicts both a greater ability to remember the location of objects and also sharper acuity in estimation. Thus, in order to plausibly account for number psychophysics, people's observed ability to detect and store the locations of objects has to be consistent with their observed estimation acuity. To test the model's predictions about both spatial memory and numerosity perception, we ran two experiments: a change-localization task to probe participants' memory for the locations of objects; and a numerical estimation task. In both experiments, we manipulated the amount of information available to participants by varying the exposure time of the presented objects. To preview our results, we find that participants' ability to remember the locations of objects — both for different exposure times and for different numbers of objects present — is entirely consistent with the observed psychophysics of number under analogous conditions.

Model

Setup

The model aims to capture how an idealized, information-limited perceptual system would perform if its only aim was to accurately store the presence or absence of objects in various locations. Although this formalizes the idea of object memory—not specifically numerical estimation—its output nonetheless yields psychophysical properties seen in number. For a given, observed scene containing objects s, we will consider the probability distribution $Q(s' \mid s)$, giving the system's belief that s' was observed instead of s. We analytically derive an optimal form of Q, by specifying three components: (i) a prior distribution representing how likely the model is to encounter a given scene *a priori*, (ii) a loss function representing how good or bad a given representation of the scene is, and (iii) 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 $Q(\cdot \mid s)$, corresponding to the optimal perceptual system. This process 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" $Q(\cdot \mid s)$ is not derived from Bayes rule, but rather from minimizing the loss function (ii) subject to an information bound (iii).

Figure 1 illustrates the basic setup, assuming for the sake of clarity that there are only 4 possible object locations (or pixels). When a person sees a particular scene, they encode a probability distribution over each possible possible arrangement of objects, which is a weighted combination of a prior for small numbers and the how well the representation matches their observation (akin to a likelihood). This probability distribution over objects 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 aren't). 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 $P(s) \propto 1/\left(n^2 \cdot \binom{M}{n}\right)$.

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}. Below we will 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'. We will assume that the loss function $\mathcal{L}(s,s')$

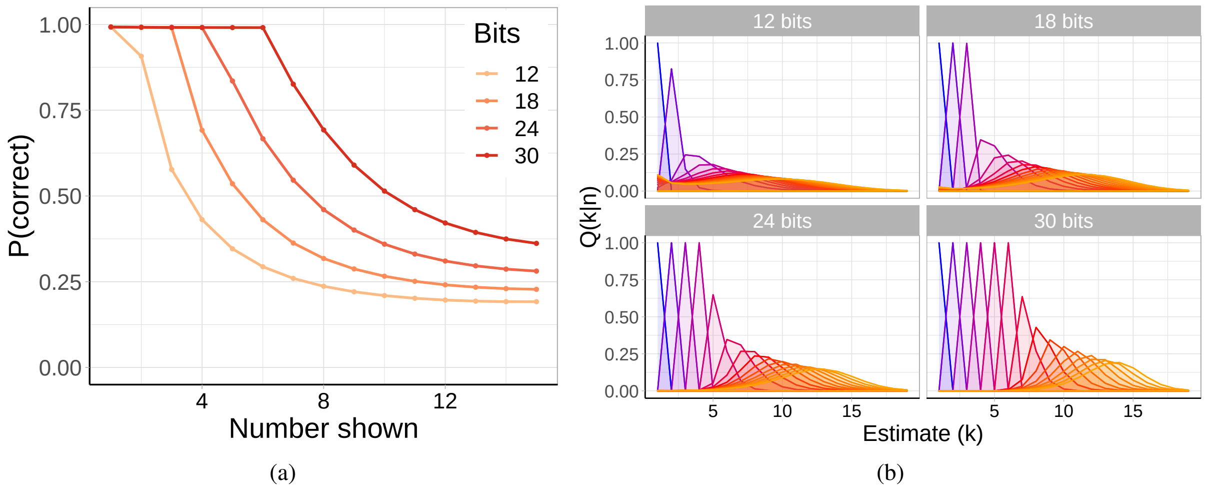


Figure 2: a) The model's predicted accuracy in a change detection task at information bounds ranging from 12-30 bits, assuming a 7x7 grid size and loss function parameter $\alpha = 1/3$ (as derived from model fitting). b) The implied psychophysics of number from the model of spatial memory. Each line shows posterior beliefs ($Q(k|n)$) over estimates (k) given numbers $n = 1 \ldots 15$. Each facet shows the results of the optimization at various information bounds (increasing left-to-right and up-to-down).

is proportional to some (perhaps unequal) combination of the proportion of false negatives, $P(s'_{ij} = 0 \mid s_{ij} = 1)$, and the proportion of false positives, $P(s'_{ij} = 1 \mid s_{ij} = 0)$. The reason we separate the contribution of false negatives and false positives here is simply that it is natural to think that the visual system might care about one more than the other. We can therefore write,

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

with α as a weighting parameter, where $0 \leq \alpha \leq 1$.

Given a loss function and prior, we now seek a function $Q(\cdot \mid s)$ that minimizes the expected loss between possible inputs s and representations s'. If the set of all possible matrices (for both inputs and representations) is S, we can write the expected loss as,

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

Unconstrained, the function $Q(\cdot \mid s)$ that minimizes the expected loss would of course be,

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

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 posterior distribution $Q(\cdot \mid s)$ over displays. The KL-divergence here represents the amount of information in bits needed to represent the posterior distribution $Q(\cdot \mid s)$ starting with the distribution $P(\cdot)$, which is equivalent to the total amount of information processing required. Given a KL-bound B we then have the constraint,

$$D_{KL}\left[Q(\cdot \mid s) \parallel P(\cdot)\right] = \sum_{s' \in R} Q(s' \mid s) \cdot \log \frac{Q(s' \mid s)}{P(s')} \leq B \quad \forall s \in S. \tag{4}$$

Now we have the components to set up a constrained optimization problem. We have an objective function (2) which quantifies how accurate a given function Q is at representing the world. We also have a bound on how costly any given Q is in terms of information processing (4). Using the method of Lagrange multipliers, we can derive an analytic solution to maximize accuracy while keeping the information processing below the information bound s',

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

for λ_s chosen to satisfy the bound in (4) for each scene s.

Results

We generated predictions from the model assuming a 7x7 grid of possible object locations, as will be used in the eventual experiments. We first simulated the model's predicted performance on a change detection task in which the model has to guess which location on the grid has changed — with an object either appearing or disappearing — between two subsequent presentations (see Experiment 1). Figure 2a shows the model's predicted accuracy (y-axis) on this task as a function of the number of objects in the scene (x-axis), at different information bounds (color). At each information bound, performance decreases as a function of the number of objects, 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 as the information bound increases, the model saturates in performance for small numbers, meaning it can veridically recall the scene it viewed.

As described above, the model's probability distribution over possible object arrangements can be converted into a probability distribution over the total number of objects. Figure 2b shows the implicit posterior (y-axis) over numerical estimates (x-axis) for each number 1-15 (lines), at the same information bounds given in Figure 2a. The model demonstrates many key properties of number psychophysics, most notably including a transition from subitizing to scalar variability. The exact point of transition, as well as the acuity of estimation, are both functions of the information bound. The model enforces a relationship between the ability to individuate and store the locations of objects in space with the ability to estimate the numerosity of the set. In this way, manipulating the information bound — e.g. by varying the presentation time of a set of objects — should allow us to test whether this relationship actually holds. That is, the psychophysics of number derived from the model are consistent with only a particular level of ability at encoding the locations of objects.

Experiment 1

The goal of Experiment 1 is to determine how much spatial information people represent about objects in a scene, which will allow us to generate predictions about number psychophysics. To do this, we constructed a change detection task in which items flashed on a screen, disappeared, and then re-appeared with a single modification. Participants' goal was to guess what changed from the first to second presentation — either which object is new or which object wasn't there previously. Each participant completed 90 trials. For half the trials, an object always appeared and for the other half an object always disappeared (split by first half of trials one way, second half the other); participants knew whether an object would be appearing or disappearing. For the sake of simplicity, we restricted the space of locations to a 7 x 7 grid with defaultly white cells, some of which (between 1 and 15) were filled gray on each trial. Finally, in order to manipulate the amount of available visual information, we varied the exposure duration (50ms, 150ms, or 450ms).

Methods

Participants We recruited 40 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 pre-screening questions. Each participant who completed the task received compensation of $3.

Materials The experiment was designed in JavaScript using the psiTurk framework (Gureckis et al., 2016). There were 49 grid cells (7 x 7), with each grid cell $35px^2$ and an equal margin separating the cells. Unfilled grid cells were white and filled grid cells were gray with hex color #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 $455px^2$ to cover the entire grid.

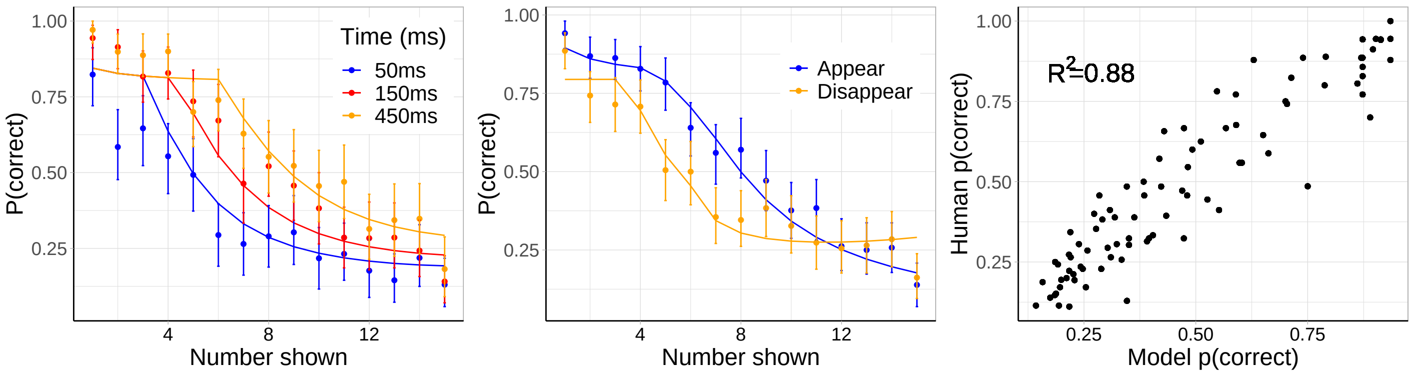
Design There were four within-subject variables manipulated in the study: the number of cells filled (1-15); the exposure time of the displayed pattern (50ms, 150ms, 450ms); and the direction of the changed cell from the first to second presentation (white-to-gray or gray-to-white). Each three-tuple of number, time, and direction was shown exactly once, for a total of 15 x 3 x 2 = 90 trials. The initial direction of 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, i.e. 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 experiment — the white-to-gray section and gray-to-white section — started with 3 practice trials. Participants were informed 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 1000 ms, followed by the grid with some cells filled in (50-450ms) 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.

Results

Our primary interest in this experiment is determining how accurate people's spatial memory for objects is as a function of the total number of objects presented and the exposure duration. Crucially, we want to test whether the model we proposed captures the key trends and provides an overall good fit to the data. To do this, we first fit model parameters to the experimental data. We assumed that the information bound changes as a function of time according to a power law $B = a \cdot t^k$, where s and k are free parameters and t is exposure time in seconds. The third free 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).

The Maximum Likelihood Estimates (MLE) for the parameters were, $a = 35.1$, $k = 0.21$, $p_g = 0.16$, and $\alpha = 0.34$ with an overall log likelihood of -1,719,. This entails information bounds of 18.7, 23.6, and 29.7 bits at 50ms, 150ms, and 450ms, 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 3a 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.



(a) (b) (c)

Figure 3: In (a) and (b) model predictions are shown as lines and human data from the change-localization task are shown as points with bootstrapped 95% confidence intervals. (a) Accuracy (y-axis) in the change-localization task as a function of the number of grid cells filled (x-axis) at each exposure duration (color). (b) 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 (color). (c) The correlation between model predictions (x-axis) and average human performance on each trial type (y-axis), i.e. trials grouped by numerosity and duration.

The model captures the effect of decreasing performance with larger quantities and increasing performance as a function of exposure duration quite well. Figure 3b 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 were fixed to 0.5, and in fact higher values of α exaggerate rather than reduce the gap between "appear" and "disapear" trials. Finally, as shown in Figure 3c, the correlation between model predictions and human accuracy across trials grouped by numerosity and exposure duration was 0.93 ($R^2 = 0.88$), indicating a good fit to the data.

Experiment 2

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. To do this, we ran a number estimation task with a design matched to Experiment 1. The procedure and display was 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. 42 adult participants from Prolific again completed 90 trials, with each number (1-15) paired with duration (50ms, 150ms, 450ms) displayed twice.

Results

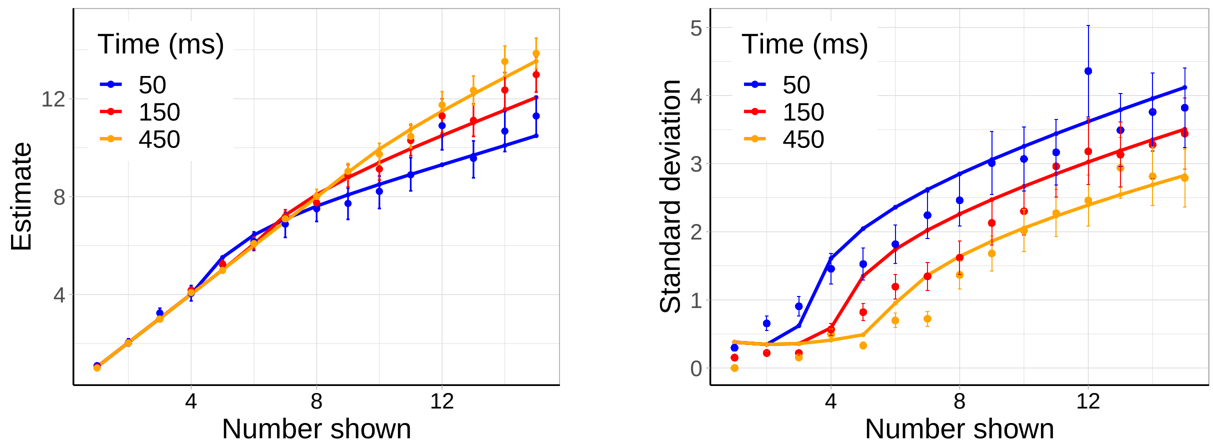
We fit the same parameters in the model with the estimation data as with the change detection task. The overall log likelihood was -6,658 and the MLE parameters were $a = 32.6$, $k = 0.21$, $p_g = 0.01$ and $\alpha = 0.32$. The implied information bounds are therefore 17.4, 21.9, and 27.6 bits at 50ms, 150ms, and 450ms respectively. This is slightly lower than the estimates derived from the change-localization task data, but the differences at each exposure duration are small. The resulting psychophysical curves from the model (lines), along with the data from the experiment (points and error-bars), are shown in Figure 4. 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. One notable difference between the inferred parameters between the two tasks is the guessing rate, which is much lower than in the change-localization task.

We compared the fit of the model to a standard psychophysical model of numerical estimation as well as a modified one that accounts for the effects of time. In the first, 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 (called a "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 MLE w fit in the non time-varying version was 0.25, with log likelihood -6,983. The MLE w_0 was -1.10 and w_t was -1.61, giving w's of 0.31, 0.26, and 0.16 at 50ms, 150ms, and 450ms respectively, and had log likelihood -6,738. The Weber models thus did not fit as well as our model, with AIC differences of over 10 (644 and 156).

Discussion

We have shown that number psychophysics emerge naturally from a model only explicitly optimized to detect and remember the locations of objects. This has some surprising implications. First, because the model accounts for subitizing as well as large number estimation, it demonstrates how a single mechanism might give rise to the observed discontinuous psychophysics. Second, it implies that the psychophysics of large number estimation — widely accepted to be a form of "gist" perception without object-level representations — may in fact arise from the process of individuating objects and tracking their locations. Finally, though 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.

There has been a long-running debate about whether performance in the subitizing range actually just reflects the approximate number system, since even if Weber's law applied



(a) (b)

Figure 4: Model predictions (lines) and data from the estimation experiment (points and 95% CI). Number is shown on the x-axis and each line represents a different exposure duration. On the y-axis: (a) mean estimates; and (b) the standard deviation of estimates.

to the small-number range, very few errors would be expected (Dehaene & Changeux, 1993; Gallistel & Gelman, 1991). The consensus view now is that the small-number range really is privileged, after carefully controlled studies found that estimation in the 1-4 range is easier than estimation of the deciles 10-40 (Revkin et al., 2008) — these have a matched ratio, and thus should be equally difficult according to Weber's law. Somewhat ironically, however, our theory posits that approximate numerical estimation is really an extension of subitizing, rather than the other way around. The difference between exact and inexact estimation, on our account, is whether the amount of information needed to represent an arrangement of objects falls above or below a capacity bound.

We previously showed that a single system optimized to estimate numerosities can explain the discontinuity in estimation ability at four as well as other aspects of number psychophysics (Cheyette & Piantadosi, 2020). However, because that model did not account for how scenes are actually encoded by vision, it made the rather unrealistic assumption that small and large numerosities are equally easy to process and that the difference in estimation precision arises purely from the fact that people need to represent small numbers more frequently (Dehaene & Mehler, 1992; Piantadosi & Cantlon, 2017). The current model, on the other hand, has the major advantage of accounting for each stage of the process, from perceptual encoding to numerical representation. So unlike the previous model, the current model predicts that large numerosities are fundamentally harder to encode with high fidelity because there are more ways to arrange many objects in space.

It is worth noting that 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 might seem to contradict predictions of our theory. 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.

Finally, we highlight two 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; Meck & Church, 1983; Plaisier et al., 2009). Though the model we presented here was designed to deal with spatial rather than temporal integration, we believe similar principles of information processing apply and hence the methods used in this paper could be adapted to capture (e.g.) the processing of auditory sequences. The other main limitation is our use of simplifying assumptions to model spatial memory — specifically, in discretizing the space so coarsely and in assuming objects to be identical. The model would thus likely 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. Gebuis et al., 2016; Gebuis & Reynvoet, 2012; Lourenco, 2015; Lourenco & Longo, 2010, 2011; Sokolowski et al., 2017). In fact, the methods we employed in this paper may be useful to understanding some of these effects: since 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

Anobile, G., Burr, D. C., Iaia, M., Marinelli, C. V., Angelelli, P., & Turi, M. (2018). Independent adaptation mechanisms for numerosity and size perception provide evidence against a common sense of magnitude. *Scientific reports*, *8*(1), 1–12.

Barth, H., Kanwisher, N., & Spelke, E. (2003). The construction of large number representations in adults. *Cognition*, *86*(3), 201–221.

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.

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

Cheyette, S. J., & Piantadosi, S. T. (2019). A primarily serial, foveal accumulator underlies approximate numerical estimation. *Proceedings of the National Academy of Sciences*, *116*(36), 17729–17734.

Cheyette, S. J., & Piantadosi, S. T. (2020). A unified account of numerosity perception. *Nature Human Behaviour*, *4*(12), 1265–1272.

Dehaene, S. (2011). *The number sense: How the mind creates mathematics*. OUP USA.

Dehaene, S., & Changeux, J.-P. (1993). Development of elementary numerical abilities: A neuronal model. *Journal of cognitive neuroscience*, *5*(4), 390–407.

Dehaene, S., & Mehler, J. (1992). Cross-linguistic regularities in the frequency of number words. *Cognition*, *43*(1), 1–29.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in cognitive sciences*, *8*(7), 307–314.

Gallistel, C., & Gelman, R. (1991). *Subitizing: The preverbal counting process*. Erlbaum Hillsdale, NJ.

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.

Gebuis, T., & Reynvoet, B. (2012). The interplay between nonsymbolic number and its continuous visual properties. *Journal of Experimental Psychology: General*, *141*(4), 642.

Green, C. S., & Bavelier, D. (2003). Action video game modifies visual selective attention. *Nature*, *423*(6939), 534.

Green, C. S., & Bavelier, D. (2006). Enumeration versus multiple object tracking: The case of action video game players. *Cognition*, *101*(1), 217–245.

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.

Inglis, M., & Gilmore, C. (2013). Sampling from the mental number line: How are approximate number system representations formed? *Cognition*, *129*(1), 63–69.

Jevons, W. S. (1871). The power of numerical discrimination. *Nature*, *3*, 281–282.

Kaufman, E. L., Lord, M. W., Reese, T. W., & Volkmann, J. (1949). The discrimination of visual number. *The American journal of psychology*, *62*(4), 498–525.

Lourenco, S. F. (2015). On the relation between numerical and non-numerical magnitudes: Evidence for a general magnitude system. *Mathematical cognition and learning* (pp. 145–174). Elsevier.

Lourenco, S. F., & Longo, M. R. (2010). General magnitude representation in human infants. *Psychological Science*, *21*(6), 873–881.

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

Mandler, G., & Shebo, B. J. (1982). Subitizing: An analysis of its component processes. *Journal of Experimental Psychology: General*, *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), 320.

Nieder, A., & Dehaene, S. (2009). Representation of number in the brain. *Annual review of neuroscience*, *32*, 185–208.

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.

Piantadosi, S. T., & Cantlon, J. F. (2017). True numerical cognition in the wild. *Psychological science*, *28*(4), 462–469.

Piazza, M., Fumarola, A., Chinello, A., & Melcher, D. (2011). Subitizing reflects visuo-spatial object individuation capacity. *Cognition*, *121*(1), 147–153.

Pica, P., Lemer, C., Izard, V., & Dehaene, S. (2004). Exact and approximate arithmetic in an amazonian indigene group. *Science*, *306*(5695), 499–503.

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.

Revkin, S. K., Piazza, M., Izard, V., Cohen, L., & Dehaene, S. (2008). Does subitizing reflect numerical estimation? *Psychological science*, *19*(6), 607–614.

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.

Stoianov, I., & Zorzi, M. (2012). Emergence of a'visual number sense'in hierarchical generative models. *Nature neuroscience*, *15*(2), 194–196.

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), 1–13.

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), e12940.