nature human behaviour

ARTICLES

https://doi.org/10.1038/s41562-020-00946-0

A unified account of numerosity perception

Samuel J. Cheyette ✉ and Steven T. Piantadosi

People can identify the number of objects in sets of four or fewer items with near-perfect accuracy but exhibit linearly increasing error for larger sets. Some researchers have taken this discontinuity as evidence of two distinct representational systems. Here, we present a mathematical derivation showing that this behaviour is an optimal representation of cardinalities under a limited informational capacity, indicating that this behaviour can emerge from a single system. Our derivation predicts how the amount of information accessible to viewers should influence the perception of quantity for both large and small sets. In a series of four preregistered experiments ($N = 100$ each), we varied the amount of information accessible to participants in number estimation. We find tight alignment between the model and human performance for both small and large quantities, implicating efficient representation as the common origin behind key phenomena of human and animal numerical cognition.

People estimate small numerosities much more rapidly and accurately than large numerosities[1–3], suggesting that we possess two separate representational systems[4,5]: a precise small-number system, which allows for the rapid identification of quantities up to around four objects with little error[1–4,6]; and an imprecise large-number system, where the standard deviation of estimates increases linearly with numerosity[5,7–10]. This hallmark of large-number estimation is known as scalar variability and can be found in many species across the animal kingdom[10–19]. However, the reason why two qualitatively different patterns of representation would arise in evolution remains obscure. Here we show that the distinct behaviour on small and large numerosities is actually expected from a single system that optimally represents quantity under a resource constraint.

Building on recent information-theoretic approaches to visual perception[20–23] and studies showing the adaptation of perceptual systems to environmental statistics[24–27], we assume that the goal of a numerical processing system is to minimize estimation error. We further assume that there is a time-dependent constraint on the numerical system's ability to process information. Under these assumptions, we present a derivation that recovers the core properties of number psychophysics, including (1) nearly exact representations for small sets[3,4,8,28], (2) scalar variability in estimation for larger numbers[5,10], (3) an underestimation bias[2,29] that diminishes with exposure time[30], (4) large-number estimation acuity that is modulated by time[30,31] and display contrast, (5) a subitizing range that is moderated by time[2] and contrast[32] and (6) roughly normally shaped response distributions for estimation[7,33]. Beyond these general properties, we test the quantitative predictions of the model about how subitizing range, estimation acuity and response distribution shape should change as functions of the amount of information perceptually available. Our results show a close agreement between human participants and bounded-optimal numerosity perception.

Results

Model set-up and assumptions. The consensus among cognitive psychologists is that at least two different systems support numerical cognition, giving rise to veridical representations of small numerosities and approximate representations of large numerosities. However, an alternative possibility is that different performance characteristics on large and small numbers result from a single psychophysical function, which itself reflects a trade-off between the benefits of veridical perception and the costs of processing sensory input. To intuitively understand this alternative, note first that most decisions that depend on numerosity involve only a small number of objects. In fact, the 'need probability'[34] of number—how often a numerosity n is encountered and represented—robustly follows a $P(n) \propto 1/n^2$ law. Empirically, the need probability is reflected in both the frequency of number words[35,36] and how often numerosities are encountered and used for decision-making in the wild[18]. This means, for instance, that we should expect that organisms need to represent seven about $1/7^2 = 1/49$th as often as they need to represent one. Efficient representational systems will take advantage of this non-uniformity and be better at representing the more frequently encountered numerosities. Second, universally in information theory, rare events require more bits of information to represent or communicate[37,38], meaning that high and low numbers will naturally place differing information processing demands in virtue of their different probabilities. Third, any organism will have a finite amount of information processing capability. This is a physical necessity and a consequence of limited perceptual systems: the amount of internal precision reserved for representations should not in general exceed the amount of information provided by perception[39].

Taken together, these facts mean that we should expect different behaviour from high and low numbers since they differ in probability; and moreover, we might expect a relatively sharp behavioural discontinuity between them if we assume a hard bound on information processing ability, with low numbers operating below the bound and high numbers operating above (and indeed, what is considered low versus high is determined by the information processing bound). We formalize these intuitions by applying standard measures from information theory and analytically computing the optimal representation given an information processing bound. These standard assumptions give rise to the details of number psychophysics as previously determined in behavioural experiments. As we show, the representation that minimizes mean squared error subject to a bounded information capacity transitions from exactness to approximation above and below the capacity bound, even though what is being optimized is a single objective function, itself representing a single system.

Consider a psychophysical function Q that maps from an observed quantity to a subjective estimate. Specifically, let $Q(k|n)$ give the probability that an observed numerosity n is represented

Department of Psychology, UC Berkeley, Berkeley, CA, USA. ✉e-mail: sam_cheyette@berkeley.edu

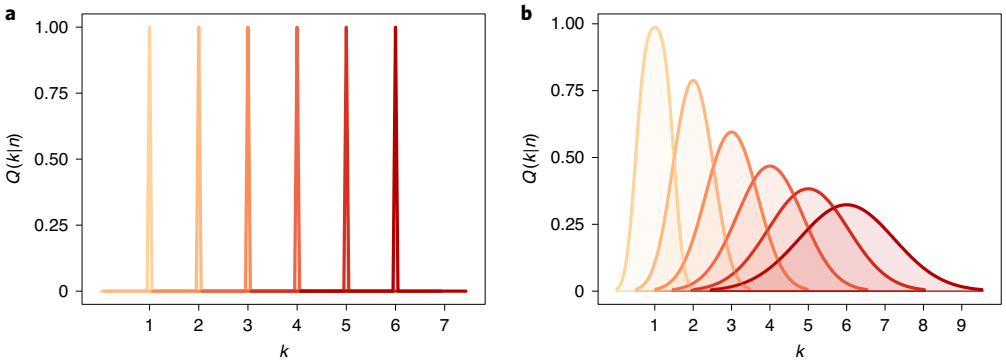

Fig. 1 | Response distributions for two possible forms of *Q*, with probabilities of estimates for numerosities 1–6. a, The form of a precise estimation system. **b**, The form of a scale variable estimation system. The numerosities are indicated by different colours.

internally with quantity k. Thus, maximally precise, veridical representations have the form

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

In general, any Q that puts a high probability on k close to n will have low error rates. Models of large-number estimation typically assume that estimates are drawn from

$$Q(k|n) \sim \text{Gaussian}(n, w\,n), \quad (2)$$

for some constant w, corresponding to scalar variability (a linear increase in the standard deviation of Q with n). The response distributions for numerosities 1–6 under these two possible forms of Q are shown in Fig. 1.

In principle, many forms of Q are logically possible, including, for example, agents that precisely represent numbers in some intermediate range or that fail completely above a given cardinality. However, we will show that the optimal Q transitions from exact solutions (as in Fig. 1a) to scalar variability (as in Fig. 1b) under some basic assumptions. First, we assume that $Q(k|n)$ is chosen to minimize the expected squared error between an input n and its representation k:

$$\mathbb{E}\left[(n-k)^2\right] = \sum\nolimits_n P(n) \sum\nolimits_k Q(k|n)(n-k)^2. \quad (3)$$

Here, $P(n)$ denotes the need probability of number, which follows a $P(n) \propto 1/n^2$ power law. Note, however, that this particular power law is not necessary to recover the key properties of the model—other need distributions exhibit similar behaviour (Supplementary Fig. 2).

If organisms had unlimited neural resources at their disposal, then the optimal Q would be given in equation (1)—that is, they would perfectly encode the numerosity of every set. But neural resources are not unlimited. Just as scientists do not usually attain measurements to more than a few digits of precision, an organism's information processing systems cannot extract arbitrary amounts of information from the world. We can formalize this constraint using a fundamental information-theoretic measure called Kullback–Leibler divergence (KL divergence)[40]. KL divergence intuitively measures how far one distribution differs from another in terms of bits of information. For instance, two overlapping distributions will have a small KL divergence, and two distributions that put most of their probability masses on different outcomes will have a high KL divergence. For us, KL divergence quantifies how many bits of information it takes to represent the distribution $Q(\cdot|n)$ starting with the distribution $P(\cdot)$, or equivalently how much information processing an organism must do to change its beliefs from $P(\cdot)$ to $Q(\cdot|n)$. It is natural, therefore, to assume that organisms with limited information processing ability will only be able to form $Q(\cdot|n)$ that are boundedly far away from $P(\cdot)$ as measured by KL divergence. In general, this bound should depend on the amount of time that an organism has to process a stimulus, since perceptual systems provide a limited bandwidth. Specifically, we assume that perception extracts information linearly in time at rate R until an overall capacity bound B is reached. Using $D_{\text{KL}}[Q(\cdot|n) \parallel P(\cdot)]$ to denote the KL divergence between $P(\cdot)$ and any hypothetical $Q(\cdot|n)$, the definition of KL divergence therefore yields the bound

$$D_{\text{KL}}[Q(\cdot|n) \parallel P(\cdot)] = \sum\nolimits_k Q(k|n) \log \tfrac{Q(k|n)}{P(k)} \leq \min(B, R\,t) \qquad \forall n. \quad (4)$$

To summarize, we are seeking a function $Q(k|n)$ that gives the probability that an organism represents n with an internal quantity k. Equation (3) defines an objective function indicating how accurate any hypothesized Q is in terms of representing the world. Equation (4) indicates how costly any hypothesized Q is in terms of information processing. Standard methods in mathematical analysis can directly derive the Q that optimizes equation (3) subject to the bound in equation (4). This is an optimization problem that can be solved (Methods) using the method of Lagrange multipliers to yield an exact analytical solution:

$$Q(k|n) \propto P(k) \exp\left(-\frac{P(n)}{\lambda_n}(n-k)^2\right) \quad (5)$$

for λ_n chosen to satisfy the bound in equation (4). This solution has a form of a weighted Gaussian with variance $\lambda_n/2P(n)$, though in our formulation this distribution is discretized. Note that the Euler–Lagrange equations of the calculus of variations can derive an analogous equation for continuous Q.

Figure 2 shows the value of $Q(\cdot|n)$ across possible numerical estimates k and the presented numerosity n, for various information capacity bounds B (faceted). The derived equation captures the following properties commonly reported in the literature on the psychophysics of number: (1) the estimation error is almost zero for small sets because they are high probability in $P(n)$ and thus require little information to specify exactly; (2) large sets exhibit scalar variability since the Gaussian component of equation (5) has a standard deviation proportional to $1/\sqrt{P(n)} \propto n$ for need distribution $P(n) \propto 1/n^2$; (3) there is an underestimation bias at low information bounds (such as two bits) due to the skew caused by the $P(k)$ term; (4) the estimation acuity (the standard deviation of $Q(k|n)$) varies with the information bound and thus presentation time; (5) the subitizing range varies with the information bound; and (6) the

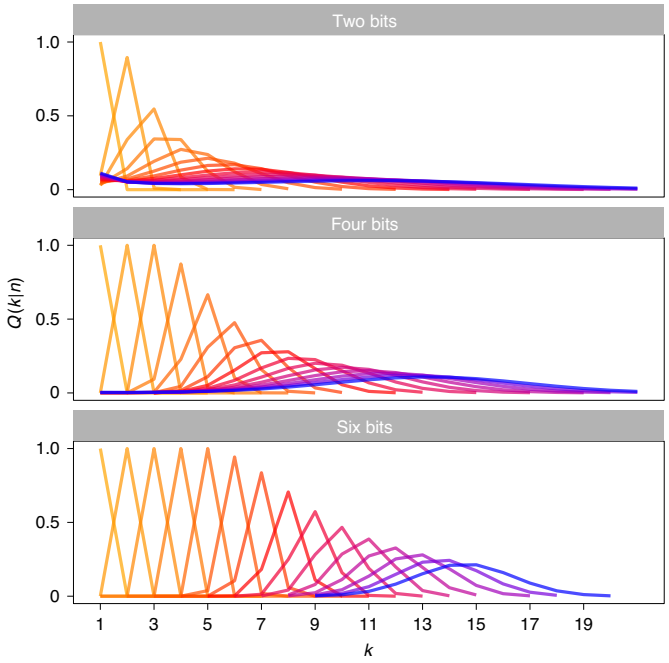

Fig. 2 | The model's posterior probability over numerosities, when shown 1 to 15 objects. The top panel shows the predictions when the model has two bits of information, the middle panel shows the predictions when it has four bits, and the bottom panel shows the predictions when it has six bits. The numerosities are indicated by different colours.

response distributions for large numerosities are roughly normally distributed, as a result of the form of equation (5).

It is important to emphasize that the roughly Gaussian tuning curves, exact representations for small sets and scalar variability are not built in as representational assumptions, but rather arise solely as a solution to the above optimization problem. The model does not even assume that $Q(k|n)$ is centred on n, and, in fact, this property only approximately holds. Note, though, that while this model shares many properties with existing psychophysical theories, equation (5) is neither an exact system nor merely an implementation of Weber's law. Instead, this equation recovers the expected behaviour of both systems in specific regimes.

Experiment. The model makes testable predictions about how estimation acuity, subitizing range and underestimation bias should depend on the amount of information available to participants. We evaluated these predictions against human behaviour in a preregistered online numerical estimation experiment (Methods). On each trial, between 1 and 15 dots were flashed, followed by a noise mask. The participants were then presented with a text box in which they typed their guess of how many dots were displayed. There were four between-participant experiments ($N = 110$ per experiment), which reflect different ways of manipulating available information (variable exposure time versus display contrast) and different ways of controlling non-numerical properties of the stimuli (the average dot size, surface area or density of the dots). Following the preregistration plan, we removed the 10 participants with the highest mean absolute error, leaving 100 participants per condition to exclude participants who weren't paying attention.

We first varied the presentation time of the dot arrays[2] (in this case holding the mean dot size constant). Varying the exposure time affects the time t in equation (4)—longer presentation times allow more information to be gathered, until the bound B is met. The dots were presented for 40 ms, 80 ms, 160 ms, 320 ms or 640 ms. We fit B and R using a hierarchical Bayesian model, which partially pools parameter estimates across participants[41] with individual participant effects on each parameter. Additionally, to account for the effects of inattention, we fit a 'guessing' parameter G, which assumes that on G proportion of trials, participants had zero bits of information about the display, meaning their estimate was effectively a random sample from their prior. Group-level parameters were given (improper) flat priors with standard deviations constrained to be positive; participant-level parameters were drawn from a normal distribution centred on the group-level mean.

The inferred group-level information rate was 48.6 bits per second (s.d. = 10.8), and the group-level bound was 4.2 bits (s.d. = 0.6), which corresponds to a subitizing range of about four. The information bound also translates to an average coefficient of variation of 0.17 for numerosities above the subitizing range. The minimum information bound (2.8 bits) corresponds to a subitizing range of merely two. This was about half of the highest information bound (5.7 bits), which corresponds to a subitizing range of five and nearly six. The inferred guessing rate was 0.05 (s.d. = 0.05), meaning that estimates in about 1 of 20 trials were probably the result of inattention. Note that some of the observed variability across participants is probably due to differences in the display, which were not tightly controlled, as is the nature of online experiments. Another point of caution is that the inferred rates and bounds would probably be lower had we not excluded the 10% of participants with the highest mean absolute error.

Figure 3a–d shows model posterior predictive fits including participant effects and human data for absolute estimation error, mean estimates and the shape of the response distributions. Critically, Fig. 3a shows that the model predicts that the error of $Q(\cdot|n)$ should also vary with presentation time, an effect found in human behaviour in Fig. 3b. Zero estimation error is found for low numbers—subitizing—in both the model and human participants at long display times. However, error increases even for small quantities at short presentation times both for the model and for human participants, reverting instead to scalar variability (a linear relationship) when the amount of available information is low. This is because less information in the input reduces the allowable KL divergence in equation (4), which forces the model to begin to approximate lower numerosities—even those in the typical subitizing range. Thus, in both people and the model, subitizing is not driven by a fixed object capacity, but rather flexibly responds to the amount of information that is visually available.

Figure 3c shows that the model predicts an underestimation bias in mean responses that diminishes at longer exposures, which is also found in human behaviour in Fig. 3d. Note that even at the shortest durations, the estimates are not random—the mean estimates still monotonically increase with the number shown in both the model and people. As predicted by the model, participants' mean estimates become increasingly unbiased at longer durations, such that the average estimate converges on the veridical number after around 160 ms. This plot shows that the model is less gradiently sensitive to time than people are, and this is probably due to our assumption of strictly linear accumulation in equation (4). Figure 3e shows the shapes of the model and human response distributions for $n = 3$, 6, 9. These make it clear that it is not just the means and standard deviations which match closely, but rather the shape of the entire distribution derived in equation (5) (which was also preregistered).

One popular alternative to a two-systems theory is that number representations are scale variable even throughout the subitizing range[9,42–45]: the error in this range under scalar variability may be small enough to yield essentially perfect accuracy. We first compared the performance of the model with an implementation of this theory, which assumes that a participant's estimate of a number n is drawn from Gaussian(n, $w \cdot n$), where w is a constant fit for each participant. To compare models, we use the Akaike information

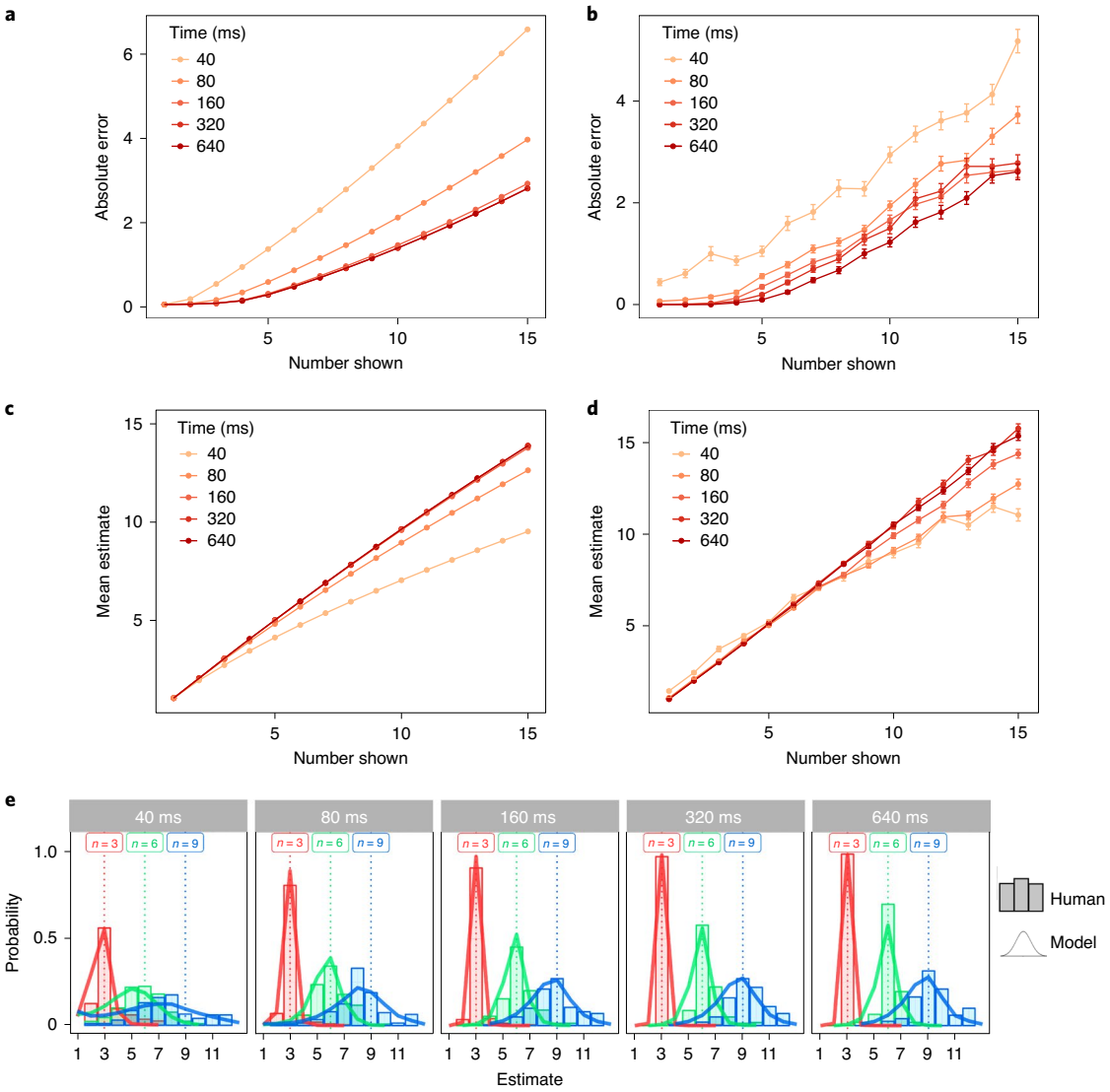

Fig. 3 | Model posterior predictive fits including participant effects and human data for absolute estimation error, mean estimates and the shape of the response distributions. a,**b**, Model predictions (**a**) and human data (**b**) for the absolute error of the estimates as a function of the number displayed and the time for the experiment with variable duration and size-controlled stimuli ($N = 100$ participants). **c**,**d**, Model predictions (**c**) and human data (**d**) for mean estimates. **e**, The probability of numeric responses over presentation times for $n = 3$, $n = 6$ and $n = 9$. Bars are shown for the human data, and lines are shown for the model predictions; all error bars represent s.e.m.

criterion (AIC), which gives better (lower) scores for models that fit data well and have few free parameters. Using maximum likelihood fits for each model, the difference in AIC scores was 3,076 in support of our model, where a difference over 10 is considered strong evidence[46]. We then fit a Weber model with linear time effects, which had an AIC difference of 768 in favour of our model. Together, these results provide strong evidence that human behaviour cannot be explained by assuming only scalar variability, or even with ad hoc modifications to scalar variability that allow acuity to vary with time and display contrast.

While the model assumes a prior $P(k) \propto 1/k^{\alpha}$ for $\alpha = 2$, the model's qualitative behaviour is robust to changes in α and can be fit with participant effects to yield similar results (Supplementary Fig. 2).

Replications. Following our preregistration plan, we sought to replicate these effects under different conditions. First, to ensure that the participants are actually using number rather than a correlated dimension, we had two groups of participants perform the same task as above but with either the total surface area or the average density of the dots controlled. Second, because other manipulations of information should have similar effects as time, we varied the display contrast[32] of the dot arrays, which affects the rate R at which information about numerosity could be extracted from the scene. In the variable-contrast experiment, the colours of the dots varied between the background (grey) and pitch black, by Weber contrasts of 10%, 20%, 40%, 80% and 160%, at a constant presentation time of 200 ms.

The inferred group-level rates and bounds were similar to those in first experiment, similarly corresponding to an average subitizing range of about four. As shown in Fig. 4a,c,e, the participants tended to underestimate larger numbers for short exposure times and low levels of contrast, matching the predictions of the model (for example, Fig. 3a). Likewise, Fig. 4b,d,f shows that in each experiment, absolute error is scale variable at low levels of informa-

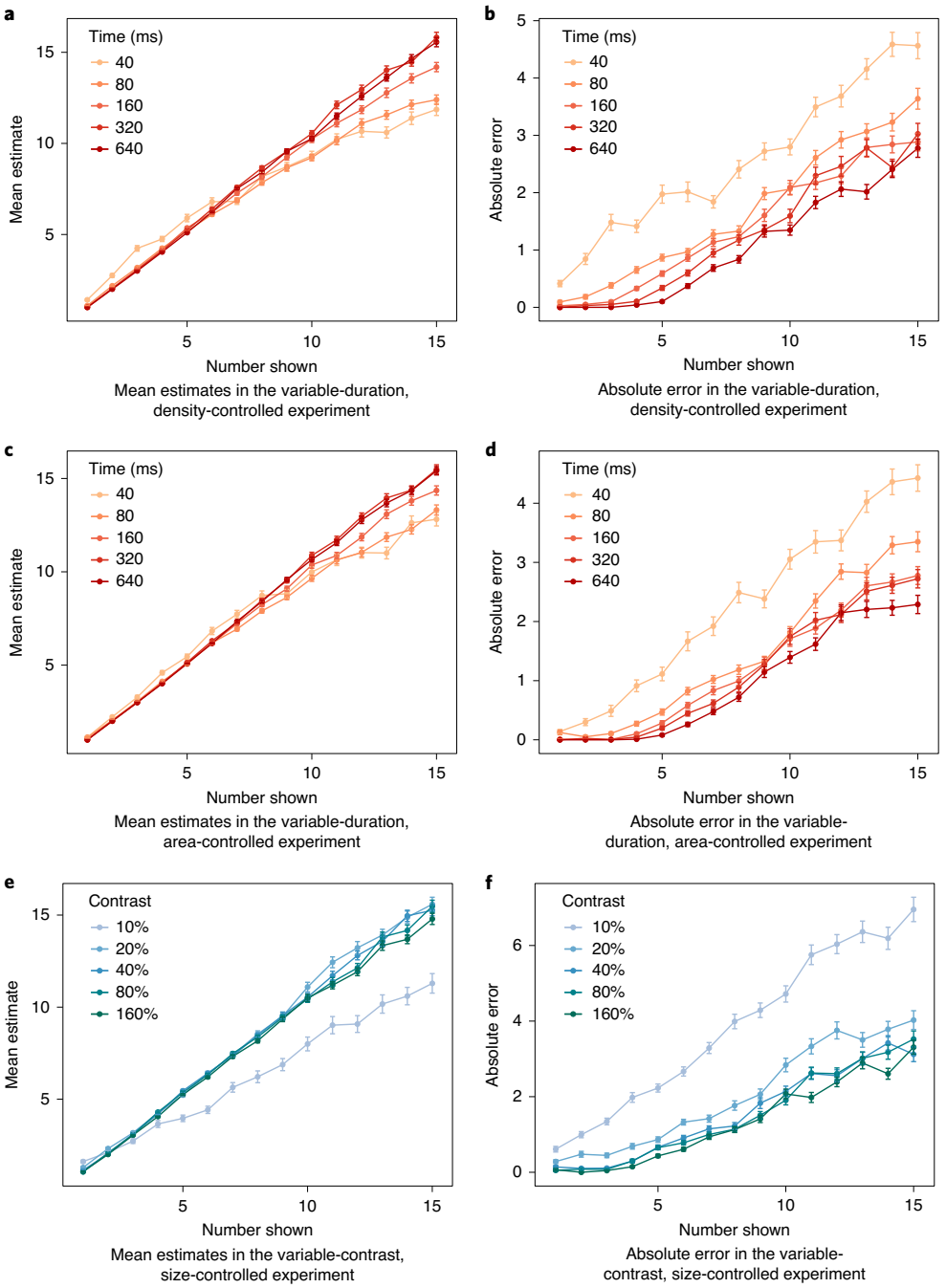

Fig. 4 | Mean estimates and absolute error of the estimates as a function of number shown in the three replication experiments. a,**c**,**e**, Mean estimates. **b**,**d**,**f**, Absolute errors of the estimates. The experiments (N = 100 each) were: variable duration with the average density controlled (**a**,**b**), variable duration with the total surface of the dots controlled (**c**,**d**) and variable display contrast with the average dot size controlled (**e**,**f**). All error bars represent s.e.m. Compare with the model in Fig. 2a,c.

tion and then becomes precise for small numbers at higher levels of information. The model provides a significantly better fit to the data from each experiment than the classic psychophysical model and its time-varying extension described above (Supplementary Information).

Discussion

Empirical studies dating back more than a century have charted many robust characteristics of numerosity perception in humans and other animals. However, most of these properties are treated as separate phenomena without a common explanation. For instance, the finding that people are able to exactly represent small sets[1,3,4,8,28] and show scalar variability in estimation for larger sets[5,10] has been explained in terms of two different representational systems[4,5]. The tendency to underestimate larger quantities[1,2] has been explained in terms of a miscalibration of response scales[29]. The sensitivity of numerical acuity to display time[2,30,31] seemingly requires ad hoc modifications to processing theories. Our derivation, however, shows that these phenomena—underestimation, distinctive behaviour on large and small sets, sensitivity to timing and contrast,

and even the shape of response distributions—can be explained as natural consequences of optimal representation under a resource constraint.

The sensitivity of numerosity judgements to certain non-numeric properties of the visual scene, such as object spacing[47] and arrangement[2,48], also fit naturally in this framework if they are considered as manipulations of information in the visual scene. For instance, regularly spaced objects appear more numerous than randomly spaced objects[48]. Likewise, objects with similar orientations appear more numerous than objects with randomly distributed orientations[49]. These effects are predicted under our model since regularities should decrease the information processing demands on the visual system.

An information-theoretic approach connects number psychophysics to the broader emerging picture of visual working memory. Contrary to a once dominant conception of visual working memory as discrete and 'slot-like'[50,51], recent behavioural and neural evidence suggests instead that visual memory flexibly allocates limited resources in a continuous manner[22,23,52–54]. Like such accounts, our model assumes that bits of information are the common currency that limit numerosity perception[39]. While others have hypothesized that subitizing is driven by a capacity limit[45], no work has formally derived how such a limit gives rise to the psychophysics of both subitizing and estimation.

Prior accounts of numerosity perception have also not explained why infants[55], some primates[18,42] and other animals[56,57] may have a smaller subitizing range than human adults. A two-systems theory would require a separate small-number system to suddenly arise in either evolution or development. However, the model we describe suggests a simple alternative: infants and many animals may have a lower visual memory capacity[58], leading the model to predict numerical approximation and scalar variability even throughout the small number range. Conversely, chimpanzees may have a subitizing range up to four or five[59], exceeding that of humans, because they have a greater visual memory capacity[60]. Similarly, an information-theoretic perspective predicts that the point at which a person transitions from subitizing to estimation should depend on their visual memory capacity, which it does[44,61,62].

More generally, this work highlights that behavioural discontinuities are not always good markers of distinct systems. Discontinuities often arise in biology when single systems face constraints—for instance, when an animal's gait varies discontinuously with its speed[63] or a neuron spikes when its input exceeds a threshold. Our results illustrate that the optimization of a single objective function may in fact show starkly different behaviour above and below a capacity bound, thus providing a resource-rational[64] account of qualitatively different patterns of numerical perception.

In sum, the theory we present relies on combining an a priori biological consideration (bounded informational capacity) with an environmental input distribution $P(n)$ and analytically computing the optimal internal representation. The resulting representational system replicates all of the standard properties of number psychophysics and explains them with a simple, resource-rational model. Our experiment has also shown that human numerical cognition quantitatively tracks this bounded optimal solution as the amount of information available varies, a fact not explainable in existing psychophysical theories. Together, these results suggest that the core properties of numerical cognition arose as efficient solutions to the problem of representing the world with finite cognitive and neural resources.

Methods

We preregistered the experiment and analysis with the Open Science Foundation on 30 October 2019. The preregistration can be found at https://osf.io/svcy5/. The experiments were approved by the University of California, Berkeley Institutional Review Board and comply with all relevant ethical regulations. Informed consent was obtained from all participants before beginning the study.

Using Lagrange multipliers to find the optimal Q. We find a form of $Q(k|n)$ chosen to minimize the expected squared error between an input n and its representation k:

$$\mathbb{E}\left[(n-k)^2\right] = \sum_{n=1}^{N} P(n) \sum_{k=1}^{N} Q(k|n)(n-k)^2 \tag{6}$$

where we have assumed an arbitrary upper bound on n of N. Here $P(n)$ is the prior on how often a number n is encountered. Assuming R is the information rate and B is the maximum allowable information, we optimize equation (6) subject to time-dependent bounds on KL divergence between Q and P:

$$D_{\mathrm{KL}}[Q(\cdot|n) \parallel P(\cdot)] = \sum_{k=1}^{N} Q(k|n) \log \frac{Q(k|n)}{P(k)} \leq \min(B, R\,t) \qquad \forall n. \tag{7}$$

Since Q is a distribution, we also have a constraint that $\sum_k Q(k|n) = 1$ for all n.

To apply the method of Lagrange multipliers, we encode the objective function and constraints into a single equation:

$$\begin{aligned}\mathcal{F}[Q(k|n)] = \quad & \sum_{n=1}^{N} P(n) \sum_{k=1}^{N} Q(k|n)(n-k)^2 \\ & + \sum_{n=1}^{N} \lambda_n \left(\min(B, Rt) - \sum_{k=1}^{N} Q(k|n) \log \tfrac{Q(k|n)}{P(k)} \right) \\ & + \sum_{n=1}^{N} \gamma_i \left(1 - \sum_{k=1}^{N} Q(k|n)\right).\end{aligned}$$

We then solve for the zeroes of the derivative of $\mathcal{F}$ with respect to $Q(k|n)$ (that is, treating $Q(k|n)$ as a separate variable for each n and k). These zeroes occur when

$$P(n)(n-k)^2 + \lambda_n \left(1 + \log \frac{\mathrm{Q(k|n)}}{\mathrm{P(k)}}\right) + \gamma_n = 0 \tag{8}$$

or

$$Q(k|n) \propto P(k) \exp\left(-\frac{P(n)}{\lambda_n}(n-k)^2\right). \tag{9}$$

Here, λ_n is chosen to satisfy the bound in equation (7).

We solve for λ_n using numerical methods. Specifically, given a bound, rate and time, we used gradient descent to find λ_n that allows the maximum $D_{\mathrm{KL}}[Q(\cdot|n) \parallel P(\cdot)]$ that satisfies the constraints. This optimizer was run for 5,000 steps for each λ_n for all numbers up to 100, which was sufficient to find KL divergences within 0.0001 bits of the bound.

Participants. We recruited 440 US adults from Amazon Mechanical Turk to participate in the experiments (110 per experiment), and they were paid US$2.50. The sample sizes were chosen on the basis of what we believed would provide reliable measures, given pilot data from the task; these sample sizes are larger than is typical for human psychophysics studies. 233 participants were female, and the mean age was 35.4 (s.d. = 12.2). We allowed only Mechanical Turk users who had above 95% acceptance rates for their work to participate. Following our preregistration plan, we removed the ten participants in each experiment whose mean absolute error was highest, leaving $N = 100$ per experiment. The data collection and analysis were not performed blind to the condition of the experiment.

Design. There were four between-participant experiments, which differed along two dimensions: the way the available information was manipulated (duration or contrast) and the way non-numerical properties of the stimuli were controlled. The two ways of manipulating the available information were varying the duration of presentation and the contrast of the dots with the background. The three non-numerical stimuli controls were the average dot size, the total surface area and the average density. Table 1 lists the pairs of these variables and controls that comprise the four experiments. The participants were randomly assigned to one of the four experimental conditions.

The experimental window was fixed to 500 × 500 pixels in any browser. However, because this was an online experiment, there were probably a range of monitor sizes and screen resolutions. We had access only to data on any browser size changes in pixels, and so we can confirm only that all browsers allowed the participants to see the full experiment, but not the physical size of the display. There was a range of window sizes, from 820 × 524 pixels to 2,560 × 1,349 pixels. The median width was 1,280 pixels, and the median height was 768 pixels.

Table 1 | The manipulated variable and how the stimuli were controlled for each of the four experiments

Experiment	Variable	Controlled
Experiment 1	Duration	Dot size
Experiment 2	Duration	Density
Experiment 3	Duration	Surface area
Experiment 4	Contrast	Dot size

The dots were presented in a 200-pixel radius around the centre of the screen. In experiments 1, 2 and 4, the average radius of the dots was fixed to 5 pixels, and each individual dot's diameter could vary between 4 and 6 pixels. In experiment 3, the total surface area of the dots was held constant by dividing the average dot size by the square root of the number of dots displayed, starting with a radius of 20 pixels for a single dot. In experiment 2, the density was fixed at approximately 900 pixels per dot, whereas in experiments 1, 3 and 4 the dots were allowed anywhere.

In each experiment, the number of dots shown was always between 1 and 15, inclusive. The participants saw each cardinality in this range twice within each of the five exposure times or contrasts (depending on the experiment). This means that, in total, the participants each completed 150 trials. The order of the stimuli was randomized over number–duration (or number–contrast) pairs.

In the variable-duration experiments, the dots were presented for 40 ms, 80 ms, 160 ms, 320 ms or 640 ms. The background was grey (hex value #B4B4B4), which was the same in the contrast experiment. The dots were darker grey, with a constant Weber contrast of 200%. In the variable-contrast experiment, the colours of the dots varied between the contrast of the background (grey) and pitch black, by Weber contrasts of 10%, 20%, 40%, 80% and 160%. This is, equivalently, black dots (hex value #000000) with opacities 4%, 8%, 16%, 32% and 64% on the grey background. The exposure duration in the variable-contrast experiment was constant at 200 ms. The noise mask covered the entire experimental window with dense, multicoloured static for 250 ms. We note that the precise display times reported here may be approximate because this study was conducted online and there may have been, for example, variation in the refresh rates of computer monitors across participants.

Procedure. After providing consent and reading the instructions, the participants were taken to the main experiment. On every trial in each experiment, a fixation cross was displayed for 750 ms, after which a number of dots were flashed on the screen. A noise mask was then applied to the screen for 250 ms, and the participants were presented with a text box in which they typed their guess of how many dots were displayed. No feedback was given. The participants were given the opportunity to take a break every ten trials.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The anonymized data from the experiments have been posted at the Open Science Foundation at https://osf.io/svcy5/.

Code availability

The code for the model can be found at https://github.com/samcheyette/info_theory_number.

Received: 19 November 2019; Accepted: 7 August 2020;
Published online: 14 September 2020

References

1. Jevons, W. S. The power of numerical discrimination. *Nature* **3**, 281–282 (1871).
2. Mandler, G. & Shebo, B. J. Subitizing: an analysis of its component processes. *J. Exp. Psychol. Gen.* **111**, 1–22 (1982).
3. Revkin, S. K., Piazza, M., Izard, V., Cohen, L. & Dehaene, S. Does subitizing reflect numerical estimation? *Psychol. Sci.* **19**, 607–614 (2008).
4. Feigenson, L., Dehaene, S. & Spelke, E. Core systems of number. *Trends Cogn. Sci.* **8**, 307–314 (2004).
5. Dehaene, S. *The Number Sense: How the Mind Creates Mathematics* (Oxford Univ. Press, 2011).
6. Kaufman, E. L., Lord, M. W., Reese, T. W. & Volkmann, J. The discrimination of visual number. *Am. J. Psychol.* **62**, 498–525 (1949).
7. Pica, P., Lemer, C., Izard, V. & Dehaene, S. Exact and approximate arithmetic in an Amazonian indigene group. *Science* **306**, 499–503 (2004).
8. Burr, D. C., Turi, M. & Anobile, G. Subitizing but not estimation of numerosity requires attentional resources. *J. Vis.* **10**, 20 (2010).
9. Gallistel, C. R. & Gelman, R. Preverbal and verbal counting and computation. *Cognition* **44**, 43–74 (1992).
10. Xu, F. & Spelke, E. S. Large number discrimination in 6-month-old infants. *Cognition* **74**, B1–B11 (2000).
11. Platt, J. R. & Johnson, D. M. Localization of position within a homogeneous behavior chain: effects of error contingencies. *Learn. Motiv.* **2**, 386–414 (1971).
12. Meck, W. H. & Church, R. M. A mode control model of counting and timing processes. *J. Exp. Psychol. Anim. Behav. Process.* **9**, 320–334 (1983).
13. Gallistel, C. R. *The Organization of Learning* (MIT Press, 1990).
14. Cantlon, J. F. & Brannon, E. M. Basic math in monkeys and college students. *PLoS Biol.* **5**, e328 (2007).
15. Cantlon, J. F. Math, monkeys, and the developing brain. *Proc. Natl Acad. Sci. USA* **109**, 10725–10732 (2012).
16. Yang, T.-I. & Chiao, C.-C. Number sense and state-dependent valuation in cuttlefish. *Proc. R. Soc. B* **283**, 20161379 (2016).
17. Uller, C., Jaeger, R., Guidry, G. & Martin, C. Salamanders (*Plethodon cinereus*) go for more: rudiments of number in an amphibian. *Anim. Cogn.* **6**, 105–112 (2003).
18. Piantadosi, S. T. & Cantlon, J. F. True numerical cognition in the wild. *Psychol. Sci.* **28**, 462–469 (2017).
19. McComb, K., Packer, C. & Pusey, A. Roaring and numerical assessment in contests between groups of female lions, *Panthera leo*. *Anim. Behav.* **47**, 379–387 (1994).
20. Sims, C. R. Rate–distortion theory and human perception. *Cognition* **152**, 181–198 (2016).
21. Sims, C. R., Jacobs, R. A. & Knill, D. C. An ideal observer analysis of visual working memory. *Psychol. Rev.* **119**, 807–830 (2012).
22. Brady, T. F., Störmer, V. S. & Alvarez, G. A. Working memory is not fixed-capacity: more active storage capacity for real-world objects than for simple stimuli. *Proc. Natl Acad. Sci. USA* **113**, 7459–7464 (2016).
23. Brady, T. F. & Tenenbaum, J. B. A probabilistic model of visual working memory: incorporating higher order regularities into working memory capacity estimates. *Psychol. Rev.* **120**, 85–109 (2013).
24. Olshausen, B. A. & Field, D. J. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* **381**, 607–609 (1996).
25. Simoncelli, E. P. & Olshausen, B. A. Natural image statistics and neural representation. *Annu. Rev. Neurosci.* **24**, 1193–1216 (2001).
26. Olshausen, B. A. & Field, D. J. Sparse coding of sensory inputs. *Curr. Opin. Neurobiol.* **14**, 481–487 (2004).
27. Geisler, W. S. Contributions of ideal observer theory to vision research. *Vis. Res.* **51**, 771–781 (2011).
28. Choo, H. & Franconeri, S. Enumeration of small collections violates Weber's law. *Psychon. Bull. Rev.* **21**, 93–99 (2014).
29. Izard, V. & Dehaene, S. Calibrating the mental number line. *Cognition* **106**, 1221–1247 (2008).
30. Cheyette, S. J. & Piantadosi, S. T. A primarily serial, foveal accumulator underlies approximate numerical estimation. *Proc. Natl Acad. Sci. USA* **116**, 17729–17734 (2019).
31. Inglis, M. & Gilmore, C. Sampling from the mental number line: how are approximate number system representations formed? *Cognition* **129**, 63–69 (2013).
32. Melcher, D. & Piazza, M. The role of attentional priority and saliency in determining capacity limits in enumeration and visual working memory. *PLoS ONE* **6**, e29296 (2011).
33. Nieder, A. & Dehaene, S. Representation of number in the brain. *Annu. Rev. Neurosci.* **32**, 185–208 (2009).
34. Anderson, J. R. & Schooler, L. J. Reflections of the environment in memory. *Psychol. Sci.* **2**, 396–408 (1991).
35. Dehaene, S. & Mehler, J. Cross-linguistic regularities in the frequency of number words. *Cognition* **43**, 1–29 (1992).
36. Piantadosi, S. T. A rational analysis of the approximate number system. *Psychon. Bull. Rev.* **23**, 877–886 (2016).
37. Stone, J. V. *Principles of Neural Information Theory* (Sebtel, 2018).
38. Shannon, C. E. A mathematical theory of communication. *Bell Syst. Tech. J.* **27**, 379–423 (1948).
39. Gallistel, C. R. Finding numbers in the brain. *Phil. Trans. R. Soc. B* **373**, 20170119 (2018).
40. Cover, T. M. & Thomas, J. A. *Elements of Information Theory* (John Wiley & Sons, 2012).
41. Gelman, A. & Hill, J. *Data Analysis Using Regression and Multilevel/Hierarchical Models* (Cambridge Univ. Press, 2006).
42. Barnard, A. M. et al. Inherently analog quantity representations in olive baboons (*Papio anubis*). *Front. Psychol.* **4**, 253 (2013).
43. Gallistel, C. & Gelman, R. in *Memories, Thoughts, and Emotions: Essays in Honor of George Mandler* (eds Kessen, W., Ortony, A. & Kraik, F.) 65–81 (Psychology Press, 1991).
44. Piazza, M., Fumarola, A., Chinello, A. & Melcher, D. Subitizing reflects visuo-spatial object individuation capacity. *Cognition* **121**, 147–153 (2011).
45. Trick, L. M. & Pylyshyn, Z. W. Why are small and large numbers enumerated differently? A limited-capacity preattentive stage in vision. *Psychol. Rev.* **101**, 80–102 (1994).
46. Anderson, D. & Burnham, K. *Model Selection and Multi-model Inference* 2nd edn (Springer, 2004).
47. Atkinson, J., Campbell, F. W. & Francis, M. R. The magic number 4 ± 0: a new look at visual numerosity judgements. *Perception* **5**, 327–334 (1976).
48. Ginsburg, N. Effect of item arrangement on perceived numerosity: randomness vs regularity. *Percept. Mot. Skills* **43**, 663–668 (1976).
49. DeWind, N. K., Bonner, M. F. & Brannon, E. M. Similarly oriented objects appear more numerous. *J. Vis.* **20**, 4 (2020).

50. Luck, S. J. & Vogel, E. K. The capacity of visual working memory for features and conjunctions. *Nature* **390**, 279–281 (1997).
51. Awh, E., Barton, B. & Vogel, E. K. Visual working memory represents a fixed number of items regardless of complexity. *Psychol. Sci.* **18**, 622–628 (2007).
52. Ma, W. J., Husain, M. & Bays, P. M. Changing concepts of working memory. *Nat. Neurosci.* **17**, 347–356 (2014).
53. Keshvari, S., Van den Berg, R. & Ma, W. J. No evidence for an item limit in change detection. *PLoS Comput. Biol.* **9**, e1002927 (2013).
54. Van den Berg, R., Shin, H., Chou, W.-C., George, R. & Ma, W. J. Variability in encoding precision accounts for visual short-term memory limitations. *Proc. Natl Acad. Sci. USA* **109**, 8780–8785 (2012).
55. Starr, A., Libertus, M. E. & Brannon, E. M. Infants show ratio-dependent number discrimination regardless of set size. *Infancy* **18**, 927–941 (2013).
56. Agrillo, C., Petrazzini, M. E. M. & Bisazza, A. Numerical acuity of fish is improved in the presence of moving targets, but only in the subitizing range. *Anim. Cogn.* **17**, 307–316 (2014).
57. Petrazzini, M. E. M., Mantese, F. & Prato-Previde, E. Food quantity discrimination in puppies (*Canis lupus familiaris*). *Anim. Cogn.* **23**, 703–710 (2020).
58. Elmore, L. C. et al. Visual short-term memory compared in rhesus monkeys and humans. *Curr. Biol.* **21**, 975–979 (2011).
59. Tomonaga, M. & Matsuzawa, T. Enumeration of briefly presented items by the chimpanzee (*Pan troglodytes*) and humans (*Homo sapiens*). *Anim. Learn. Behav.* **30**, 143–157 (2002).
60. Inoue, S. & Matsuzawa, T. Working memory of numerals in chimpanzees. *Curr. Biol.* **17**, R1004–R1005 (2007).
61. Green, C. S. & Bavelier, D. Action video game modifies visual selective attention. *Nature* **423**, 534–537 (2003).
62. Green, C. S. & Bavelier, D. Enumeration versus multiple object tracking: the case of action video game players. *Cognition* **101**, 217–245 (2006).
63. Alexander, R. M. The gaits of bipedal and quadrupedal animals. *Int. J. Rob. Res.* **3**, 49–59 (1984).
64. Griffiths, T. L., Lieder, F. & Goodman, N. D. Rational use of cognitive resources: levels of analysis between the computational and the algorithmic. *Top. Cogn. Sci.* **7**, 217–229 (2015).

Acknowledgements

We thank F. Callaway, J. Cantlon and E. Gibson for providing feedback on an earlier draft of this paper. This work was supported by grants no. 1760874 and no. 2000759 from the National Science Foundation, Division of Research on Learning (to S.T.P.) and award no. 1R01HD085996 from the Eunice Kennedy Shriver National Institute of Child Health & Human Development (NICHD) at the National Institutes of Health (to S.T.P. and J. Cantlon). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

S.J.C. and S.T.P. derived and implemented the model. S.J.C. and S.T.P. designed the experiment. S.J.C. implemented the experiment and analysed the data. S.J.C. and S.T.P. wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41562-020-00946-0.

Correspondence and requests for materials should be addressed to S.J.C.

Peer review information Primary handling editor: Aisha Bradshaw.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

nature research

Corresponding author(s): Samuel Cheyette

Last updated by author(s): Jul 22, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	The experiment was implemented using Psiturk (version 2.3.6), with custom code in HTML and Javascript. We collected data using Amazon Mechanical Turk.
Data analysis	The model was implemented using Python 2.7.15, with custom functions for computing the model's predictions at different amounts of information and for fitting the data. We used R 3.4.4 for all other statistical analyses, using the lme4 package.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Anonymized data from the experiments have been posted at the Open Science Foundation at https://osf.io/svcy5/.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	*Describe how sample size was determined, detailing any statistical methods used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient.*
Data exclusions	*Describe any data exclusions. If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*
Replication	*Describe the measures taken to verify the reproducibility of the experimental findings. If all attempts at replication were successful, confirm this OR if there are any findings that were not replicated or cannot be reproduced, note this and describe why.*
Randomization	*Describe how samples/organisms/participants were allocated into experimental groups. If allocation was not random, describe how covariates were controlled OR if this is not relevant to your study, explain why.*
Blinding	*Describe whether the investigators were blinded to group allocation during data collection and/or analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.*

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	We ran 4 experiments on human adult subjects, collecting and analyzing their behavioral data (their estimates of numbers at different exposure durations/levels of contrast) .
Research sample	We collected data from 440 participants (233 female, mean age 35.4 (SD=12.2)). Each participant completed 150 trials.
Sampling strategy	We ran the four experiments simultaneously, randomly assigning participants to each. We pre-determined the sample size (as noted in our pre-registration).
Data collection	Participants were linked to the experiment on Amazon Mechanical Turk's website. The only equipment was each participant's computer. The number of dots shown, the location of every dot, the amount of time the dots were shown, the contrast level of the dots, and the participants' reaction times were all recorded.
Timing	Data were collected in 8 batches (to minimize server load) over the course of 2 days: October 30th and November 2nd.
Data exclusions	We excluded the 10 participants with the highest mean absolute error in each experiment, following the pre-registratrion plan. This left us with 100 participants per experiment.
Non-participation	71 participants began but did not complete the experiment. No reasons were given.
Randomization	Participants were randomly assigned to one of the four experimental conditions (with counter-balancing to ensure correct numbers).

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	*Briefly describe the study. For quantitative data include treatment factors and interactions, design structure (e.g. factorial, nested, hierarchical), nature and number of experimental units and replicates.*
Research sample	*Describe the research sample (e.g. a group of tagged Passer domesticus, all Stenocereus thurberi within Organ Pipe Cactus National Monument), and provide a rationale for the sample choice. When relevant, describe the organism taxa, source, sex, age range and any manipulations. State what population the sample is meant to represent when applicable. For studies involving existing datasets, describe the data and its source.*
Sampling strategy	*Note the sampling procedure. Describe the statistical methods that were used to predetermine sample size OR if no sample-size*

Sampling strategy	*calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient.*
Data collection	*Describe the data collection procedure, including who recorded the data and how.*
Timing and spatial scale	*Indicate the start and stop dates of data collection, noting the frequency and periodicity of sampling and providing a rationale for these choices. If there is a gap between collection periods, state the dates for each sample cohort. Specify the spatial scale from which the data are taken*
Data exclusions	*If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*
Reproducibility	*Describe the measures taken to verify the reproducibility of experimental findings. For each experiment, note whether any attempts to repeat the experiment failed OR state that all attempts to repeat the experiment were successful.*
Randomization	*Describe how samples/organisms/participants were allocated into groups. If allocation was not random, describe how covariates were controlled. If this is not relevant to your study, explain why.*
Blinding	*Describe the extent of blinding used during data acquisition and analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.*

Did the study involve field work? ☐ Yes ☐ No

Field work, collection and transport

Field conditions	*Describe the study conditions for field work, providing relevant parameters (e.g. temperature, rainfall).*
Location	*State the location of the sampling or experiment, providing relevant parameters (e.g. latitude and longitude, elevation, water depth).*
Access and import/export	*Describe the efforts you have made to access habitats and to collect and import/export your samples in a responsible manner and in compliance with local, national and international laws, noting any permits that were obtained (give the name of the issuing authority, the date of issue, and any identifying information).*
Disturbance	*Describe any disturbance caused by the study and how it was minimized.*

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☐ Antibodies
☐	☐ Eukaryotic cell lines
☐	☐ Palaeontology
☐	☐ Animals and other organisms
☐	☐ Human research participants
☐	☐ Clinical data

Methods

n/a	Involved in the study
☐	☐ ChIP-seq
☐	☐ Flow cytometry
☐	☐ MRI-based neuroimaging

Antibodies

Antibodies used	*Describe all antibodies used in the study; as applicable, provide supplier name, catalog number, clone name, and lot number.*
Validation	*Describe the validation of each primary antibody for the species and application, noting any validation statements on the manufacturer's website, relevant citations, antibody profiles in online databases, or data provided in the manuscript.*

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	*State the source of each cell line used.*
Authentication	*Describe the authentication procedures for each cell line used OR declare that none of the cell lines used were authenticated.*
Mycoplasma contamination	*Confirm that all cell lines tested negative for mycoplasma contamination OR describe the results of the testing for*

Mycoplasma contamination	*mycoplasma contamination OR declare that the cell lines were not tested for mycoplasma contamination.*
Commonly misidentified lines (See ICLAC register)	*Name any commonly misidentified cell lines used in the study and provide a rationale for their use.*

Palaeontology

Specimen provenance	*Provide provenance information for specimens and describe permits that were obtained for the work (including the name of the issuing authority, the date of issue, and any identifying information).*
Specimen deposition	*Indicate where the specimens have been deposited to permit free access by other researchers.*
Dating methods	*If new dates are provided, describe how they were obtained (e.g. collection, storage, sample pretreatment and measurement), where they were obtained (i.e. lab name), the calibration program and the protocol for quality assurance OR state that no new dates are provided.*

☐ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	*For laboratory animals, report species, strain, sex and age OR state that the study did not involve laboratory animals.*
Wild animals	*Provide details on animals observed in or captured in the field; report species, sex and age where possible. Describe how animals were caught and transported and what happened to captive animals after the study (if killed, explain why and describe method; if released, say where and when) OR state that the study did not involve wild animals.*
Field-collected samples	*For laboratory work with field-collected samples, describe all relevant parameters such as housing, maintenance, temperature, photoperiod and end-of-experiment protocol OR state that the study did not involve samples collected from the field.*
Ethics oversight	*Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.*

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Human research participants

Policy information about studies involving human research participants

Population characteristics	All participants were U.S. adults (age 18 and over).
Recruitment	Participants were recruited on the Amazon Mechanical Turk platform, required to have high work-acceptance rates (>95%).
Ethics oversight	The study was approved by the University of California, Berkeley IRB.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	*Provide the trial registration number from ClinicalTrials.gov or an equivalent agency.*
Study protocol	*Note where the full trial protocol can be accessed OR if not available, explain why.*
Data collection	*Describe the settings and locales of data collection, noting the time periods of recruitment and data collection.*
Outcomes	*Describe how you pre-defined primary and secondary outcome measures and how you assessed these measures.*

ChIP-seq

Data deposition

☐ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☐ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links *May remain private before publication.*	*For "Initial submission" or "Revised version" documents, provide reviewer access links. For your "Final submission" document, provide a link to the deposited data.*

Files in database submission	*Provide a list of all files available in the database submission.*
Genome browser session (e.g. UCSC)	*Provide a link to an anonymized genome browser session for "Initial submission" and "Revised version" documents only, to enable peer review. Write "no longer applicable" for "Final submission" documents.*

Methodology

Replicates	*Describe the experimental replicates, specifying number, type and replicate agreement.*
Sequencing depth	*Describe the sequencing depth for each experiment, providing the total number of reads, uniquely mapped reads, length of reads and whether they were paired- or single-end.*
Antibodies	*Describe the antibodies used for the ChIP-seq experiments; as applicable, provide supplier name, catalog number, clone name, and lot number.*
Peak calling parameters	*Specify the command line program and parameters used for read mapping and peak calling, including the ChIP, control and index files used.*
Data quality	*Describe the methods used to ensure data quality in full detail, including how many peaks are at FDR 5% and above 5-fold enrichment.*
Software	*Describe the software used to collect and analyze the ChIP-seq data. For custom code that has been deposited into a community repository, provide accession details.*

Flow Cytometry

Plots

Confirm that:

☐ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☐ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☐ All plots are contour plots with outliers or pseudocolor plots.

☐ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	*Describe the sample preparation, detailing the biological source of the cells and any tissue processing steps used.*
Instrument	*Identify the instrument used for data collection, specifying make and model number.*
Software	*Describe the software used to collect and analyze the flow cytometry data. For custom code that has been deposited into a community repository, provide accession details.*
Cell population abundance	*Describe the abundance of the relevant cell populations within post-sort fractions, providing details on the purity of the samples and how it was determined.*
Gating strategy	*Describe the gating strategy used for all relevant experiments, specifying the preliminary FSC/SSC gates of the starting cell population, indicating where boundaries between "positive" and "negative" staining cell populations are defined.*

☐ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Magnetic resonance imaging

Experimental design

Design type	*Indicate task or resting state; event-related or block design.*
Design specifications	*Specify the number of blocks, trials or experimental units per session and/or subject, and specify the length of each trial or block (if trials are blocked) and interval between trials.*
Behavioral performance measures	*State number and/or type of variables recorded (e.g. correct button press, response time) and what statistics were used to establish that the subjects were performing the task as expected (e.g. mean, range, and/or standard deviation across subjects).*

Acquisition

Imaging type(s) — *Specify: functional, structural, diffusion, perfusion.*

Field strength — *Specify in Tesla*

Sequence & imaging parameters — *Specify the pulse sequence type (gradient echo, spin echo, etc.), imaging type (EPI, spiral, etc.), field of view, matrix size, slice thickness, orientation and TE/TR/flip angle.*

Area of acquisition — *State whether a whole brain scan was used OR define the area of acquisition, describing how the region was determined.*

Diffusion MRI ☐ Used ☐ Not used

Preprocessing

Preprocessing software — *Provide detail on software version and revision number and on specific parameters (model/functions, brain extraction, segmentation, smoothing kernel size, etc.).*

Normalization — *If data were normalized/standardized, describe the approach(es): specify linear or non-linear and define image types used for transformation OR indicate that data were not normalized and explain rationale for lack of normalization.*

Normalization template — *Describe the template used for normalization/transformation, specifying subject space or group standardized space (e.g. original Talairach, MNI305, ICBM152) OR indicate that the data were not normalized.*

Noise and artifact removal — *Describe your procedure(s) for artifact and structured noise removal, specifying motion parameters, tissue signals and physiological signals (heart rate, respiration).*

Volume censoring — *Define your software and/or method and criteria for volume censoring, and state the extent of such censoring.*

Statistical modeling & inference

Model type and settings — *Specify type (mass univariate, multivariate, RSA, predictive, etc.) and describe essential details of the model at the first and second levels (e.g. fixed, random or mixed effects; drift or auto-correlation).*

Effect(s) tested — *Define precise effect in terms of the task or stimulus conditions instead of psychological concepts and indicate whether ANOVA or factorial designs were used.*

Specify type of analysis: ☐ Whole brain ☐ ROI-based ☐ Both

Statistic type for inference (See Eklund et al. 2016) — *Specify voxel-wise or cluster-wise and report all relevant parameters for cluster-wise methods.*

Correction — *Describe the type of correction and how it is obtained for multiple comparisons (e.g. FWE, FDR, permutation or Monte Carlo).*

Models & analysis

n/a	Involved in the study
☐	☐ Functional and/or effective connectivity
☐	☐ Graph analysis
☐	☐ Multivariate modeling or predictive analysis

Functional and/or effective connectivity — *Report the measures of dependence used and the model details (e.g. Pearson correlation, partial correlation, mutual information).*

Graph analysis — *Report the dependent variable and connectivity measure, specifying weighted graph or binarized graph, subject- or group-level, and the global and/or node summaries used (e.g. clustering coefficient, efficiency, etc.).*

Multivariate modeling and predictive analysis — *Specify independent variables, features extraction and dimension reduction, model, training and evaluation metrics.*