Sampling in Approximate Number Perception

Emily M. Sanford (esanford@berkeley.edu) & Steven T. Piantadosi (stp@berkeley.edu)
Department of Psychology, University of California, Berkeley

Abstract

Approximate number perception is noisy, but it is unclear what kind of underlying process the noise reflects. Here we provide evidence that approximate number estimation should be thought of as a sampling procedure. We show that the the average of two approximate number estimates of the same stimulus tends to outperform either estimate alone; additionally, the average difference between the two estimates of a given number linearly increases as a function of number, consistent with Weber's law. Finally, we provide evidence that people report confidence ranges consistent with Weber's law. This suggests that they represent a distribution of possible responses even on a single trial.

Keywords: Approximate number system, psychophysics, sampling, distribution

Introduction

The ability to estimate large numbers, often called the Approximate Number Sense (ANS), is a key exemplar of quantitative representation (Dehaene, 1997; Feigenson, Dehaene, & Spelke, 2004). Sensitivity to the number of objects in an ensemble has been displayed in a variety of populations and species (Bryer et al., 2022), from fish (Agrillo, Piffer, & Bisazza, 2011; Bisazza, Tagliapietra, Bertolucci, Foà, & Agrillo, 2014) and salamanders (Krusche, Uller, & Dicke, 2010; Uller, Jaeger, Guidry, & Martin, 2003) to human infants, children and adults (Halberda & Feigenson, 2008; Izard, Sann, Spelke, & Streri, 2009; Xu & Spelke, 2000). A fair amount is known about the representation of approximate quantities in the brain and mind. One central phenomenon is *scalar variability*: estimates are imprecise, with the amount of variability (standard deviation) increasing linearly as the number increases, consistent with Weber's law (Dehaene, 2003; Halberda & Feigenson, 2008; Piazza, Izard, Pinel, Le Bihan, & Dehaene, 2004). A standard model of approximate number representation (Dehaene, 2003; Gallistel & Gelman, 2000) formalizes this by assuming that we possess Gaussian curves which sit atop a mental number line, and the degree to which the curves are activated as possible responses depends on their distance to the true value. The standard deviations of these curves increase linearly with the represented number, giving the distribution,

$$P(\text{responding k} \mid \text{shown n}) \sim Normal(n, w \cdot n)$$

This model is supported by empirical response distributions, where the spread in estimates of a given value increases as the value increases, or equivalently, numbers become harder to discriminate as they get larger (Gallistel & Gelman, 2000; Platt & Johnson, 1971). The model is also supported by neuronal evidence, from a parieto-frontal brain network—particularly the intraparietal sulcus—where, in monkeys, there are number-selective neurons, whose activation is highest for their preferred number and progressively declines with increasing numerical distance from the preferred number (Nieder & Miller, 2004; Nieder, 2005, 2016; Roitman, Brannon, & Platt, 2007). The Gaussian curves can be decoded from neuronal activity in the brains of the animals performing numerical tasks, and an analogous result has been obtained with fMRI in human subjects as well (Piazza et al., 2004).

What has not been explicitly delineated, and was often assumed but unstated in much of the previous work, is the step-by-step procedure that a subject undertakes when making a numerical estimate based on a stimulus. Here we study several of the procedural steps involved in constructing an estimate. Subject must first acquire information from the world about the perceptual stimulus. Next, they activate or generate a representation based on that information. Finally, they output a response in the form of an estimate.

In Experiment 1, we ask whether number estimation involves stochastic sampling in the statistical sense. Sampling is a process where observations are drawn from a distribution, and the likelihood of any particular observation being selected depends on the height of the distribution at that point. In number estimation, a sampling process could occur either during the perception stage (gathering information about the stimulus), or at the response stage (outputting a response based on the representation). Our preliminary goal is to establish whether *at least one* of these two stages provides the statistical benefits of a sampling procedure, as opposed to another process like maximizing. To do this, we take advantage of the fact that, when sampling from a normal distribution, the average of multiple samples will tend to be closer to the mean of the distribution than any of the samples alone (Vul & Pashler, 2008). Therefore, to test whether number estimation involves sampling (either during the perceptual or response stages), we ask subjects to provide two number estimates in response to each stimulus. If these two responses are sam-

In M. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the 45th Annual Conference of the Cognitive Science Society.*

pled from a normal distribution, then the average of those two responses will tend to be more precise than either estimate alone. If the average of two responses does *not* outperform either alone, this would indicate that no sampling is taking place. In this case, number estimation responses may instead be better described as one's "best guess" of the number of objects in the stimulus. This would mean that both the perception stages and the response stages are essentially non-random: every time a subject sees the same stimulus, they end up with (approximately) the same representation, and every time they have that representation, they provide the same response (e.g., the maximum, if the representation is distributed).

Next, we investigate what format the mental representation mediating number estimation takes. As stated previously, neural and behavioral evidence suggests that number representations have distributional properties (Dehaene, 2003; Nieder, 2005; Piazza et al., 2004). However, the distribution-like results could be consistent with either an internally represented *distribution*, or instead an internally represented *point estimate*, which becomes a normal distribution only when averaged across many trials. Some previous evidence has supported the probabilistic nature of number representations by demonstrating that people take into account the relative fidelity of different information sources (e.g., vision and audition) when making number estimates (Kanitscheider, Brown, Pouget, & Churchland, 2015). Here we ask whether these representations not only reflect one's confidence about which information source to rely upon, but also the typical psychophysical characteristics associated with number estimation behavior (i.e., scalar variability).

In Experiment 2, we ask subjects to provide intervals that they are confident contain the true number of dots in the image. If the representations underlying numerical responses are probabilistic or distributional, we would expect subjects to provide larger ranges for larger numbers, consistent with Weber's law (Gallistel & Gelman, 2000). If these representations are point estimates instead (albeit tagged with information about the relative fidelity of the input source), we would expect to find no relationship between the value being estimated and the width of the interval provided by subjects. This would indicate that the distributions associated with number responses only emerge when multiple trials are aggregated, but are not represented in the mind at the single trial level.

Experiment 1

Methods

Subjects

Subjects were online workers recruited on the platform Prolific. 200 subjects completed the experiment and were paid $1.64 for their participation.

Materials

Stimuli were randomly generated for each subject. Each stimulus consisted of between 5 and 29 light gray dots on a dark gray background. The dots could only appear at predetermined locations within a grid. The grid had 10 rows, and the number of columns varied based on the user's screen size (between 5 and 28 columns) to meet the constraint that the rows and columns be equally spaced. There were two instances of each number (5 to 29) generated, yielding 50 unique stimuli.

Procedure

The experiment was completed on subjects' own computers over the internet. Subjects provided informed consent, then were instructed to estimate the number of dots appearing on the screen. Each stimulus was presented for 250 ms, followed by a static mask displayed for 500 ms. A blank screen with a text box remained until a typed response was submitted. The experiment started with 5 practice trials with no feedback, then subjects completed 100 experimental trials, divided into two blocks with a self-paced break in between. Unbeknownst to the participants, the second block consisted entirely of repeats of the 50 unique trials shown in the first block, in a different randomized order. As a result, subjects provided two temporally separated estimates for the number of dots in each of the 50 stimuli. Subjects were debriefed, paid, and thanked for their participation. The experiment took a median time of 11:39 to complete.

Note that, because the two estimates were provided after *two separate viewings* of the stimulus, with this method we cannot distinguish whether any positive evidence of sampling we find comes from the perceptual stage or the response stage (or both).

Results

Analyses were preregistered on AsPredicted (#112280).

Data Cleaning

Following our preregistration, we excluded individual trials as outliers if the response was beyond the 5% or 95% cutoffs of responses given for each number. Of the 19,800 total trials, 1708 were removed (including all of one participant's responses), leaving 18,092 trials for analysis. Additionally, the entire dataset with no outliers removed was analyzed in the Median Absolute Deviation analysis (see below), as this analysis is robust to outliers.

The average of two guesses is better than either alone

If subjects are sampling from a Weber-like distribution when making an estimate, rather than each estimate reflecting their best guess given all available evidence from the stimulus, we would expect that the average of their two separate estimates would tend to perform better than either guess alone.

Mean Squared Error (MSE) The first way we calculated response error was using the Mean Squared Error (MSE) between responses and the true number depicted, following Vul and Pashler (2008). We separately computed the MSE for each trial across subjects for each of the two responses each

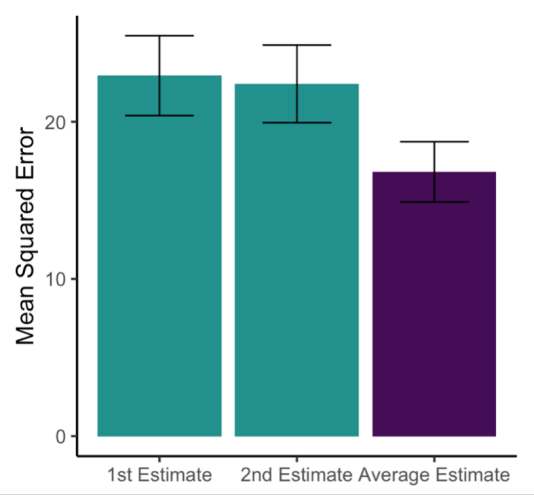

Figure 1: The MSE of the average of two estimates is lower than the MSE for either estimate alone.

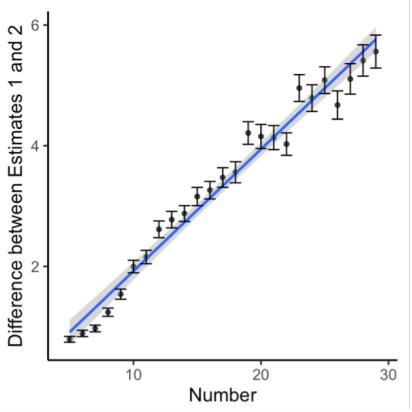

Figure 2: The average difference between two estimates for a given stimulus increases linearly as a function of the number of dots depicted in that stimulus.

subject provided, as well as for the average of each subject's two responses for that trial. We found that the average of two responses easily outperformed either of the single estimates alone (see Figure 1). Average responses (right bar) had less error than either the 1st or 2nd response alone (left and middle bars, respectively). This result was confirmed with a nonparametric Friedman test, $\chi = 75.39$, $p < .001$, Kendall's $W = .754$ (indicating a large effect size), and follow up Wilcoxon signed rank test with Bonferroni corrections indicated that the average estimate was different from both the 1st and 2nd estimates, $ps < .001$, while those estimates did not differ from each another, $p = 1$.

Median Absolute Deviation (MAD) In addition to the MSE, we also calculated the Median Absolute Deviation (MAD) of the responses, as this measure is robust to outliers. Following our pre-registration, no trials were excluded from this analysis (N = 19,800). When analyzing the entire sample with the MAD, the effect remained: the average of two estimates outperformed either single estimate alone. This was again confirmed by a Friedman test, $\chi = 16.36$, $p < .001$, Kendall's $W = .164$ (indicating a small effect size). Once again, follow up Wilcoxon signed rank test with Bonferroni corrections indicated that the average estimate was different from either the 1st or 2nd estimates, $ps < .005$, while the 1st and 2nd estimate did not differ from one another, $p = 1$.

Samples come from distributions that exhibit scalar variability

Thus far, we have demonstrated that number estimation responses show properties of stochastic sampling. Next, we asked whether the underlying distributions from which the responses are being sampled are Weber-like, as would be expected with numerical behavior (Dehaene, 2003; Piazza et al., 2004). If responses are sampled from distributions that follow Weber's law, we would expect that the difference between two estimates for a given stimulus should increase, on average, as the number of dots depicted increase – that is, they should exhibit scalar variability. We found that this was in fact the case, and the relationship was notably linear (see Figure 2). Using the package brms to run a Bayesian mixed effects linear regression with random intercepts and slopes for subjects (Bürkner, 2018), we found that there was a strong relationship between the number of dots in the stimulus and the distance between the subject's two estimates for that stimulus, $B = .21$, $95\%CI = [.19, .22]$, and this model outperformed the null model, $\Delta LOOIC = 1969$.

Psychophysical measures of performance

Finally, we also evaluated whether classic psychophysical measures of estimation performance differed between the 1st and 2nd estimate as well as their average. We fit each subject's responses with a Bayesian modeling approach in Stan (Carpenter et al., 2017). The model contained two parameters: a Weber fraction (w) and a measure of bias (b). The model assumes that estimation responses are drawn (i.e., sampled) from a Gaussian distribution with a mean of N^b and a standard deviation of $N^b * w$, where N is the true number, b corresponds to how much the representation is under ($b < 1$) or overestimated ($b > 1$) relative to the true value, and w is an index of internal variability or precision (smaller w corresponds to a more precise representation). The model was fit across subjects separately for each estimate (1st, 2nd, and average).

We found that fitted Weber fractions were reliably smaller, indicating better precision, for the average of the two estimates as compared to either estimate alone (see Figure 3). There was much less difference in bias; the bias of all three estimates hovered very near to unbiased responding at $b = 1$.

Discussion

In this experiment, we have provided evidence that the process of number estimation involves a stochastic sampling pro-

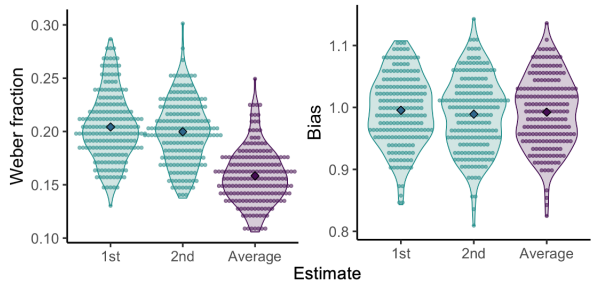

Figure 3: Model-fitted psychophysical parameters.

cedure. Different viewings of the same stimulus lead to different responses, and this difference cannot be entirely attributed to noise (e.g., noisy attempts to report the same response at different times), since the average of two responses performs better than either alone.

Experiment 2

Next, we ask whether people internally represent a distribution after viewing a stimulus for a very brief duration, or if the process of sampling leads to the representation of a point estimate. In the latter case, the resemblance of numerical responses to a distribution would emerge only by combining across many trials.

Methods

Subjects

Subjects were again recruited from Prolific; 201 people participated (all distinct from Experiment 1; although we preregistered N = 200, 1 additional subject participated due to experimenter error, and their data are nonetheless included in analyses as their inclusion did not change the results). Subjects were paid $1.70 for their participation.

Materials

Stimuli were generated using the same procedure as described in Experiment 1.

Procedure

In this experiment, instead of asking for a single estimate, we asked subjects to provide an *interval* containing the correct response. Following our consent procedure, subjects were provided with the following instructions:

> To respond, you will provide a range of intuitively plausible values that you are very confident contains the correct number. For example, if an image actually contains 15 dots, you might answer "Lower Number: 10, Higher Number: 20" if you're not so sure of the exact number. Or, you might answer "Lower Number: 15, Higher Number: 15" if you are completely certain that the answer is 15. Try to provide the smallest range that you can, while still capturing the true value inside the range.

As in Experiment 1, the stimuli appeared on the screen for 250 ms, followed by a static mask displayed for 500 ms. Then, two text boxes – one labeled "Low Number?" and one labeled "High Number?" – remained on the screen until a response was entered in both. Subjects received 5 practice trials with no feedback followed by 50 experimental trials, with a self-paced break in the middle. The experiment took a median time of 9:59 to complete.

Results

Analyses were preregistered on AsPredicted (#112280).

Data Cleaning

Of the total 10,050 trials, 21 were excluded because the range provided was negative (i.e., the "high number" was lower than the "low number;" visual inspection of these trials indicated that they were mostly typos). Following our preregistration, we excluded trials as outliers if the size of the range (high number minus low number) was beyond the 5% or 95% cutoffs of responses given for each number across subjects. With 10,029 remaining trials, this criterion removed 698, leaving 9,331 for analysis.

Interval size varies as a function of number

If the format of subjects' number representation is distributed or probabilistic (consistent with behavioral psychophysics of number estimation), range estimates would get wider as the number of dots in the stimulus increased. We found this to be robustly the case; the average ranges given for each number are displayed in Figure 4. The x-axis corresponds to the number being estimated, and the y-axis corresponds to the estimates provided. Note that only the high and low numbers were actually supplied by participants; the points in the middle represent the midpoints of those ranges. People provided increasingly wider interval estimates when trying to capture large numbers as compared to small numbers: for 5 dots, the

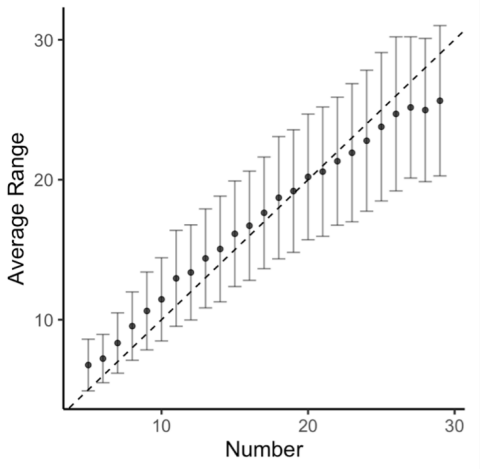

Figure 4: Average ranges provided for each number; the dashed line corresponds to perfect estimation.

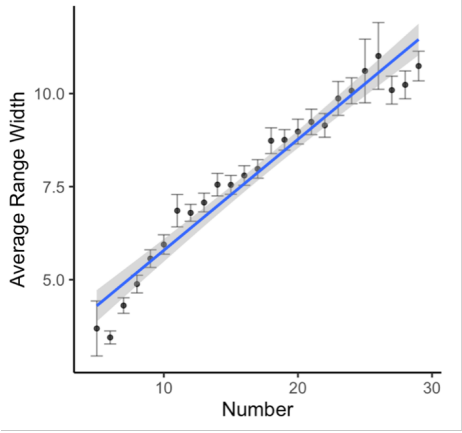

Figure 5: Number strongly predicts the average width of ranges provided by subjects.

average interval was 4.92 to 8.61 (width = 3.69), while for 29 dots, the average range was 20.28 to 31.01 (width = 10.74).

We confirmed the linear effect of number on range size using a Bayesian mixed effects linear regression with random subject effects (see Figure 5). The number of dots in the stimulus was highly predictive of range width, $B = .29$, $95\%CI = [.27, .31]$, and the model explained significant variance over the null model, $\Delta LOOIC = 5766.6$.

Discussion

In this experiment, we found that subjects have access to variability in their representations. They can explicitly indicate that they are less certain and therefore require a wider range to capture the correct value when the number is larger. This indicates that the underlying representations are in fact distributions, and not simply point estimates. However, we note that multiple samples may be used to estimate these distributions and intervals.

General Discussion

In this set of experiments, we have demonstrated that number perception reflects a stochastic sampling process. This aligns with previous work, which has argued that tasks like visually object identification similarly rely on sampling (Moreno-Bote, Knill, & Pouget, 2011). We also provided evidence that the mental representations involved in number estimation are distributed, also consistent with prior research (Kanitscheider et al., 2015).

As previously stated, with the current method we are unable to determine whether sampling is occurring at the perceptual stage, the response stage, or both. Much research supports the hypothesis that visual perception behaves much like a Bayesian sampling procedure (Kanitscheider et al., 2015; Knill & Richards, 1996; Moreno-Bote et al., 2011; Townsend, Hu, & Ashby, 1981). Other research has suggested that responding on the basis of one's own probabilistically-distributed internal representation also involves a sampling procedure, at least in the context of factual "trivia-like" questions (Vul & Pashler, 2008). If the mental representation employed during number estimation were a point estimate, it would not be possible to sample from it to provide a response. However, given that we the internal mental representations appear to be distributed, this leaves open the possibility that the sampling in Experiment 1 reflects sampling from one's internal probability distribution. Therefore, it is entirely possible that *both* stages involve sampling. We will investigate this in future research.

Another open question is how the mental representation of a distribution is stored. One possibility could be a sampler, where only the samples themselves are represented and accessible, with no explicit means to calculate probabilities (Sanborn & Chater, 2016). Another possibility is that parameters characterizing the distribution are actually calculated, such that the representation could be compressed to something like a mean and standard deviation. It is relatively simple to convert between these two formats, and they would both lead to extremely similar behaviors. Therefore, further research will be necessary to differentiate between the two formats.

Acknowledgments

This research was supported by NSF grant 2000759.

References

Agrillo, C., Piffer, L., & Bisazza, A. (2011). Number versus continuous quantity in numerosity judgments by fish. *Cognition*, *119*(2), 281–287. doi: 10.1016/j.cognition.2010.10.022

Bisazza, A., Tagliapietra, C., Bertolucci, C., Foà, A., & Agrillo, C. (2014). Non-visual numerical discrimination in a blind cavefish (Phreatichthys andruzzii). *Journal of Experimental Biology*, *217*(11), 1902–1909. doi: 10.1242/jeb.101683

Bryer, M. A., Koopman, S. E., Cantlon, J. F., Piantadosi, S. T., MacLean, E. L., Baker, J. M., … Vonk, J. (2022). The evolution of quantitative sensitivity. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *377*(1844), 20200529. doi: 10.1098/rstb.2020.0529

Bürkner, P.-C. (2018). Advanced Bayesian Multilevel Modeling with the R Package brms. *The R Journal*, *10*(1), 395. doi: 10.32614/RJ-2018-017

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., … Riddell, A. (2017). Stan : A Probabilistic Programming Language. *Journal of Statistical Software*, *76*(1). doi: 10.18637/jss.v076.i01

Dehaene, S. (1997). *The Number Sense: How the Mind Creates Mathematics*. New York, New York: Oxford University Press.

Dehaene, S. (2003). The neural basis of the Weber-Fechner law: A logarithmic mental number line. *Trends in Cognitive Sciences*, *7*(4), 145–147. doi: 10.1016/S1364-6613(03)00055-X

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

Gallistel, C. R., & Gelman, R. (2000). Non-verbal numerical cognition: From reals to integers. *Trends in Cognitive Sciences*, *4*(2), 59–65. doi: 10.1016/S1364-6613(99)01424-2

Halberda, J., & Feigenson, L. (2008). Developmental Change in the Acuity of the "Number Sense": The Approximate Number System in 3-, 4-, 5-, and 6-Year-Olds and Adults. *Developmental Psychology*, *44*(5), 1457–1465. doi: 10.1037/a0012682

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

Kanitscheider, I., Brown, A., Pouget, A., & Churchland, A. K. (2015). Multisensory decisions provide support for probabilistic number representations. *Journal of Neurophysiology*, *113*(10), 3490–3498. doi: 10.1152/jn.00787.2014

Knill, D. C., & Richards, W. (Eds.). (1996). *Perception as Bayesian Inference*. Cambridge University Press. doi: 10.1017/CBO9780511984037

Krusche, P., Uller, C., & Dicke, U. (2010). Quantity discrimination in salamanders. *Journal of Experimental Biology*, *213*(11), 1822–1828. doi: 10.1242/jeb.039297

Moreno-Bote, R., Knill, D. C., & Pouget, A. (2011). Bayesian sampling in visual perception. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(30), 12491–12496. doi: 10.1073/pnas.1101430108

Nieder, A. (2005). Counting on neurons: The neurobiology of numerical competence. *Nature Reviews Neuroscience*, *6*(3), 177–190. doi: 10.1038/nrn1626

Nieder, A. (2016). The neuronal code for number. *Nature Reviews Neuroscience*, *17*(6), 366–382. doi: 10.1038/nrn.2016.40

Nieder, A., & Miller, E. K. (2004, may). A parieto-frontal network for visual numerical information in the monkey. *Proceedings of the National Academy of Sciences*, *101*(19), 7457–7462. doi: 10.1073/pnas.0402239101

Piazza, M., Izard, V., Pinel, P., Le Bihan, D., & Dehaene, S. (2004). Tuning curves for approximate numerosity in the human intraparietal sulcus. *Neuron*, *44*(3), 547–555. doi: 10.1016/j.neuron.2004.10.014

Platt, J. R., & Johnson, D. M. (1971). Localization of position within a homogeneous behavior chain: Effects of error contingencies. *Learning and Motivation*, *2*(4), 386–414. doi: 10.1016/0023-9690(71)90020-8

Roitman, J. D., Brannon, E. M., & Platt, M. L. (2007). Monotonic coding of numerosity in macaque lateral intraparietal area. *PLoS Biology*, *5*(8), 1672–1682. doi: 10.1371/journal.pbio.0050208

Sanborn, A. N., & Chater, N. (2016). Bayesian Brains without Probabilities. *Trends in Cognitive Sciences*, *20*(12), 883–893. doi: 10.1016/j.tics.2016.10.003

Townsend, J. T., Hu, G. G., & Ashby, F. G. (1981). Perceptual sampling of orthogonal straight line features. *Psychological Research*, *43*(3), 259–275. doi: 10.1007/BF00308451

Uller, C., Jaeger, R., Guidry, G., & Martin, C. (2003). Salamanders (Plethodon cinereus) go for more: Rudiments of number in an amphibian. *Animal Cognition*, *6*(2), 105–112. doi: 10.1007/s10071-003-0167-x

Vul, E., & Pashler, H. (2008). Measuring the crowd within: Probabilistic representations within individuals: Short report. *Psychological Science*, *19*(7), 645–647. doi: 10.1111/j.1467-9280.2008.02136.x

Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition*, *74*(1), 1–11. doi: 10.1016/S0010-0277(99)00066-9