

# UC Berkeley

## UC Berkeley Previously Published Works

### Title

Successful discrimination of tiny numerical differences

### Permalink

<https://escholarship.org/uc/item/9fm5k536>

### Journal

Journal of Numerical Cognition, 9(1)

### Authors

Sanford, Emily M

Halberda, Justin

### Publication Date

2023-03-31

### DOI

10.5964/jnc.10699

### Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

## Successful Discrimination of Tiny Numerical Differences

Emily M. Sanford<sup>1</sup> , Justin Halberda<sup>1</sup> 

[1] *Department of Psychological and Brain Sciences, Johns Hopkins University, Baltimore, MD, USA.*

---

Journal of Numerical Cognition, 2023, Vol. 9(1), 196–205, <https://doi.org/10.5964/jnc.10699>

Received: 2022-11-18 • Accepted: 2023-01-13 • Published (VoR): 2023-03-31

Handling Editor: Joonkoo Park, University of Massachusetts Amherst, Amherst, MA, USA

Corresponding Author: Emily M. Sanford, who is now at the Department of Psychology, University of California, Berkeley, 2121 Berkeley Way, Berkeley, CA 94704, United States. E-mail: [esanford@berkeley.edu](mailto:esanford@berkeley.edu)

---

### Abstract

Are there some differences so small that we cannot detect them? Are some quantities so similar (e.g., the number of spots on two speckled hens) that they simply look the same to us? Although modern psychophysical theories such as Signal Detection Theory would predict that, with enough trials, even minute differences would be perceptible at an above-chance rate, this prediction has rarely been empirically tested for any psychological dimension, and never for the domain of number perception. In an experiment with over 400 adults, we find that observers can distinguish which of two collections has more dots from a brief glance. Impressively, observers performed above chance on every numerical comparison tested, even when discriminating a comparison as difficult as 50 versus 51 dots. Thus, we present empirical evidence that numerical discrimination abilities, consistent with SDT, are remarkably fine-grained.

### Keywords

approximate number sense, magnitude discrimination, psychophysics, limits, guessing

What are the limits of our perception? This question has been a central focus of the scientific study of the mind since the investigation of thresholds began with psychophysics (Fechner, 1966/1860; Weber, 1978/1834). Although Signal Detection Theory (SDT) attempted to do away with the idea of thresholds entirely (e.g., Laming, 1973), the idea that at some point differences can be so small as to be imperceptible remains a prominent concept in modern psychological theory. Indeed, we imagine that all readers share with us the intuition that two clouds of dots can become so close in number that it is impossible for a human to judge which has more after just a brief glance (e.g., two orange trees, one with 200 oranges and the other with 201). Here, we investigate whether, contrary to these intuitions, we are in fact sensitive to very small differences, and that the function governing perceptual discrimination performance is indeed smooth all the way down.

While the argument we will explore applies broadly to all magnitude discriminations, the literature on approximate number cognition is a particularly informative test case, as limits on our ability to judge number have been widely discussed. The ability to rapidly determine the numerically greater of two collections is attested across species and across human development (e.g., Feigenson et al., 2004). The core system thought to be responsible for representing non-symbolic numerical quantities is known as the Approximate Number System (ANS), and it may be linked to more advanced mathematical abilities across the lifespan (Halberda et al., 2008; Mazzocco et al., 2011; Park & Brannon, 2013; Starr et al., 2017; Wang et al., 2016, 2017; but see De Smedt et al., 2013; Norris & Castronovo, 2016; Qiu et al., 2021; Szűcs & Myers, 2017). ANS acuity is often measured using discrimination tasks, where two ensembles are presented, and the subject decides which contains more objects. With such comparisons, trial difficulty is determined by the



ratio of the compared groups (Feigenson et al., 2004). Individuals differ in how precise their numerical estimates tend to be (Halberda et al., 2008; Halberda et al., 2012), and this has been accounted for by Weber's law – a relationship between the ratio of the values to be discriminated and performance first described by Fechner in the earliest years of experimental psychology (Fechner, 1966/1860).

The idea that some differences are too small to perceive is prevalent throughout the literature on numerical cognition, as exemplified by the following quote: “The difference between eight and nine is not experienced at all, since eight and nine, like any higher successive numerical values, cannot be discriminated” (Carey, 2009, p. 295). The classification of small differences as imperceptible in numerical cognition is particularly noticeable in the developmental literature, where, due in part to practical limitations on experiment length, rather than modeling complete psychometric curves, experimenters often opt to test only a few ratios and determine which result in above-chance performance at their limited sample size. For instance, one of the earliest studies establishing numerical sensitivity in infants concluded that 6-month-olds require a 2:1 ratio in order to perceive a difference between large approximate collections (Xu & Spelke, 2000). In fact, the use of above versus at-chance performance as a distinguishing metric for performance is present widely across the number literature for a variety of populations and species (e.g., ratio 5:4 for mosquitofish, Agrillo et al., 2007; 1.7:1 for angelfish, Gómez-Laplaza & Gerlai, 2011; 5:4 for cotton-top tamarinds, Hauser et al., 2003; 2:1 for newborn infants, Izard et al., 2009; 6:4 for red-backed salamanders, Uller et al., 2003). The implication of these results, either stated or unstated by the authors, is that ratios smaller than some critical value cannot be differentiated by that population under those conditions.

However, the conclusion that these populations are *insensitive* to smaller differences is inconsistent with modern models of psychophysics, such as SDT. Instead, these models predict smoothly decreasing performance as a function of the ratio of the comparison, with no drop to “chance performance” even on extremely difficult ratios. It is an inconsistency, then, that these two conventions – results claiming to reflect “limits” on number discrimination, as well as modeling approaches informed by SDT (e.g., Halberda et al., 2008; Piazza et al., 2010; Pica et al., 2004) – coexist in the numerical cognition literature. Further, although success is *theoretically* predicted on such intuitively “impossible” trials, they have yet to be strongly *empirically tested* with ample enough sample sizes to find a significant effect, at least in numerical cognition. Here we provide this empirical test.

This last point is of particular theoretical significance, as there is precedent for revision of psychophysical theory in response to new data from edge cases: for example, previous models of magnitude perception such as Weber's law were eventually found to be inadequate to explain behavior at extremes such as extremely high and low stimulus intensities (McGill & Goldberg, 1968). And, in numerical discrimination tasks, it is conceivable that human participants may employ behavioral strategies that lead to deviations from the predictions of SDT (i.e., above chance performance on even the most difficult comparisons). One such strategy, which we here call the “Give Up” function, assumes that subjects cease to respond on the basis of their representations when comparisons become especially difficult – instead opting to respond randomly (for an example of results that appear consistent with this strategy, see Halberda & Feigenson, 2008). Therefore, even with a thorough understanding and belief in the validity of modern psychophysical laws, it is not *given* that these laws will characterize behavior all the way to the limit (i.e., equality).

Here, we empirically test whether the psychophysical theories of magnitude comparison (using number as a case study) are indeed correct even in extreme cases, despite their inconsistency with the intuitive appeal of hard limits and the convention of using such limits to characterize the acuity of perceptual systems. That is, are we above chance at numerical discrimination, even for extremely difficult comparisons? Importantly, most of the comparisons that we presented were far below the levels typically employed in numerical discrimination experiments, with the most difficult ratio being 51:50 dots. Although many authors in numerical cognition may expect the contrary, either due to a belief in true perceptual limits or an intuition that subjects would give up on subjectively “impossible” comparisons, we hypothesized that, with enough trials, people would perform above chance on even the hardest ratios.

## Method

### Subjects

We ran batches until we had at least 100 subjects for each of four conditions. A total of  $N = 412$  people completed the experiment ( $n = 110$  for stimulus set A with sequential presentation,  $n = 100$  for the stimulus set B with sequential presentation,  $n = 102$  participants for stimulus set A with simultaneous presentation, and  $n = 100$  participants for stimulus set B with simultaneous presentation; see Materials for details of the stimulus sets). Subjects were undergraduate students enrolled in a psychology course who received course credit for their online participation.

### Materials

On each trial, subjects saw two ensembles and indicated which of the two groups contained more dots. There was always a correct answer, so equal numbers of dots were never shown. Each stimulus image was displayed for 1000 ms. There were two conditions. Half of subjects saw the groups sequentially, at the center of the screen (all white dots against a gray background), with a 50 ms blank screen between ensemble presentations. The other half of subjects saw the groups simultaneously, with one group on the left (white dots) and one on the right (black dots). Following the stimuli, a response prompt (“Did the first/left or second/right image contain more dots?”) remained on the screen until a response was recorded on the subject’s keyboard.

Stimulus sets A and B were distinct sets of images generated using the same algorithm, which were both included solely for the purpose of ensuring generalizability of our results beyond one particular stimulus set. This means that there were essentially four “conditions” with approximately equal numbers of participants: stimulus set A with sequential presentation, stimulus set B with sequential presentation, stimulus set A with simultaneous presentation, and stimulus set B with simultaneous presentation.

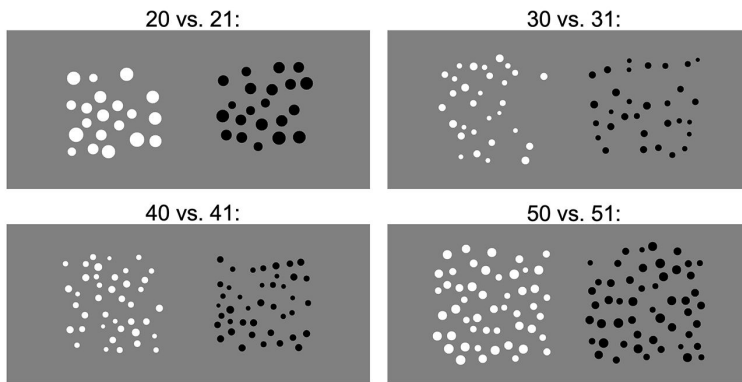
Finally, during stimulus generation, we implemented non-numerical feature controls in an attempt to focus subjects on number. Surface area and convex hull ratios on each trial were approximately equated to that trial’s number ratio, with equal numbers of congruent and incongruent trials for each feature. However, we note that our broader argument – that subjects are capable of above chance discrimination performance on extremely difficult ratios – would stand whether subjects in our task were relying on number, area, convex hull, or on any combination of features (i.e., the ratio of any feature on the 51:50 trials was equal to 51:50 – and far more difficult than the ratios typically used in number, area and convex hull experiments).

### Procedure

There were four blocks of trials, with participant-paced breaks in between each block. Each block began with practice trials on easy ratios that scaffolded the participant towards the difficult test ratio (Confidence Hysteresis; Odic et al., 2014). For every block, the practice trials consisted of 8 trials each of comparisons of 30:20, then 40:30, then 50:40, such that there were 24 practice trials per block. Following each round of practice trials, subjects responded to 56 trials of one of the difficult ratios, with randomized block order: 21:20, 31:30, 41:40, or 51:50. These ratios were selected because even the easiest (21:20) is much more difficult than ratios typically employed in numerical comparison tasks (e.g., 2:1 – 10:9; Halberda & Feigenson, 2008), and the most difficult (51:50) is subjectively extremely difficult, yielding a sense of a complete lack of confidence in pilot testing (see Figure 1). Trial order was randomized across participants within each block, and the order in which stimuli were presented (left vs. right/1<sup>st</sup> vs. 2<sup>nd</sup>) was counterbalanced, such that each option was the correct response on exactly half of the trials in that condition.

**Figure 1**

Example Stimuli From Each of the Four Hard Comparisons



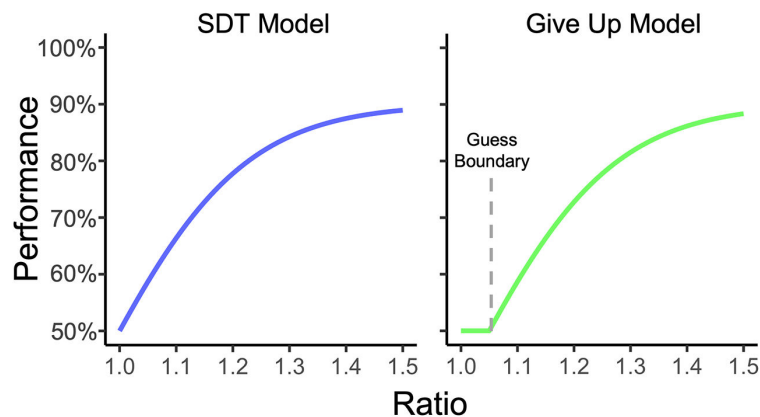
Subjects were tested online on their own personal computer during COVID, so factors like total display size and luminance were not tightly controlled – but these should have only minor effects on performance, given the previously-found reliability of internet-based psychological studies (e.g., [Germine et al., 2010](#)).

## Computational Models

We fit each subject's response data with two models (see [Figure 2](#)). The first is a modification of a psychophysical model derived from SDT used extensively in previous ANS research (e.g., [Halberda et al., 2008](#); [Piazza et al., 2010](#); [Pica et al., 2004](#)), which here we call the SDT Model. This model takes responses to trials with different ratios and fits an internal Weber fraction ( $w$ ) and guess rate ( $g$ ). The Weber fraction corresponds to the rate of expansion of the standard deviation of the ANS's Gaussian representations and captures the amount of non-overlap between the Gaussian representations of the compared numbers (e.g., [Dehaene, 2003](#)). A smaller  $w$  corresponds to more precise responding. The guess rate accounts for the amount that subjects are blindly guessing on any trial regardless of ratio. An increased  $g$  results in a lower ceiling on performance and affects performance on other ratios proportionately (i.e., it does not result in increased guessing specifically on difficult ratios, as would be seen with giving up).

**Figure 2**

Tested Psychophysical Models



The probability ( $p$ ) that a subject responds correctly (i.e., correctly identifies which of the two groups has the larger number of dots) is a function of the ratio of the numerosity of the compared ensembles,  $r = \frac{n_{larger}}{n_{smaller}}$ , according to the following function:

$$p = (1 - g) \times \left[ 1 - \frac{1}{2} \operatorname{erfc} \left( \frac{r - 1}{w\sqrt{2}\sqrt{1 + r^2}} \right) \right] + \frac{g}{2}$$

As can be seen in left panel of Figure 2, this curve is the top half of a sigmoid, and the inflection point in the sigmoid would happen at  $r = 1$  (when the two groups have equal numbers of dots). This means that the SDT model predicts that performance will be above chance for all ratios until equality ( $r = 1$ ), with sufficient sample size.

For the alternative model, which we are calling the Give Up Model, we devised a modification of the SDT model in an attempt to capture the intuition that comparisons eventually become so difficult as to be imperceptible – i.e., that they give rise to random guessing. In our Give Up Model, if the ratio was more difficult than the Guess Boundary (i.e.,  $r < 1 + \text{Guess Boundary}$ ), the probability of a correct response was 50%. If the ratio was easier than the Guess Boundary (i.e.,  $r > 1 + \text{Guess Boundary}$ ), the probability of a correct response was determined using the SDT Model for a ratio of  $r = r - \text{Guess Boundary}$  (which allows the function to be continuous; see Figure 2, right). While the specific way in which participants may “blend” their responses based on the signal (SDT) and their responses based on guesses (Give Up) near the Guess Boundary has yet to be determined, this straightforward approach will allow us to detect whether or not a Give Up function is necessary to account for human performance:

$$p = \begin{cases} (1 - g) \times \left[ 1 - \frac{1}{2} \operatorname{erfc} \left( \frac{(r - \text{Guess Boundary}) - 1}{w\sqrt{2}\sqrt{1 + (r - \text{Guess Boundary})^2}} \right) \right] + \frac{g}{2} & \text{if } r > \text{Guess Boundary} \\ .5 & \text{if } r < \text{Guess Boundary} \end{cases}$$

## Results

### Exclusions

We excluded subjects based on average accuracy across the whole experiment. Overall, subjects were correct on 62.0% of trials ( $SD = 4.4\%$ ). Subjects who performed more than three standard deviations below the mean for accuracy (that is, below 48.7% across all trials) were excluded from the analysis. This exclusion criterion resulted in the removal of 2 participants. As a result, 410 participants were included in analyses. Note that, because performance is expected to vary by ratio, this criterion leaves plenty of room for observing chance performance across the group on harder ratios.

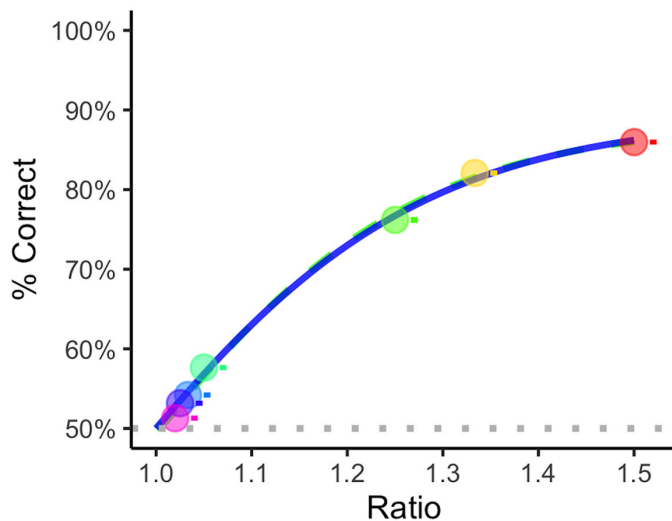
### Performance Above Chance

Our main question of interest is whether subjects can perceive the difference between groups at the most difficult numerical ratios. For each ratio, we calculated each subject’s accuracy. Then we performed a series of planned one-sample  $t$ -tests comparing average performance with chance (50%), separately for each ratio.

Subjects performed above chance for all ratios, all  $ps < .001$  (see Figure 3). For the easiest comparison in the practice trials (30:20, ratio = 1.5), subjects averaged 86.0% accuracy ( $SD = 10.8\%$ ). On the hardest comparison (51:50, ratio = 1.02), subjects averaged 51.3% accuracy ( $SD = 6.4\%$ ), which was still significantly higher than chance,  $t(409) = 4.11$ ,  $p < .001$  (pink dot in Figure 3). Average performance increased smoothly as a function of ratio.

Figure 3

Performance on the Number Comparison Task



Note. The blue line corresponds to the best-fitting group-level SDT model; The green dashed line (which is almost completely obscured by the blue line) corresponds to the best-fitting group-level Give Up Model.

To evaluate whether performance varied by condition, we ran a 2-way (stimulus set x presentation method) between-subjects ANOVA on overall accuracy. There was a marginal difference in performance between subjects who saw stimulus set A ( $M = 62.3\%$ ,  $SD = 4.3\%$ ) and stimulus set B ( $M = 61.6\%$ ,  $SD = 4.6\%$ ),  $F(1,408) = 3.18$ ,  $p = .075$ . There was a significant difference in performance by presentation method. Subjects in the sequential presentation condition had higher accuracy ( $M = 63.3\%$ ,  $SD = 4.1\%$ ) than subjects in the simultaneous presentation condition ( $M = 60.5\%$ ,  $SD = 4.3\%$ ),  $F(1,408) = 44.47$ ,  $p < .001$ . The interaction between stimulus set and presentation condition was not significant,  $p = .309$ .

Because of the difference in performance between the presentation conditions, we evaluated whether subjects were significantly above chance at each ratio separately for each presentation condition. In the sequential presentation condition, subjects were significantly above chance on every comparison including 51:50 ( $M = 52.0\%$ ,  $SD = 6.6\%$ ),  $ps < .001$ . Subjects in the simultaneous presentation condition were above chance on every comparison,  $ps < .001$ , *except* for 51:50, which was not significantly different from chance ( $M = 50.5\%$ ,  $SD = 6.0\%$ ),  $t(201) = 1.28$ ,  $p = .202$ . Note that the simultaneous presentation (1000 ms total) allowed exactly half of the amount of time allowed in the serial presentation (2000 ms total) and this, rather than any lack of ability, may be the cause of the slightly lower percent correct on these trials. More importantly, this deviation was not large enough in magnitude for a Give Up model to be preferred even for responses in this condition (see below).

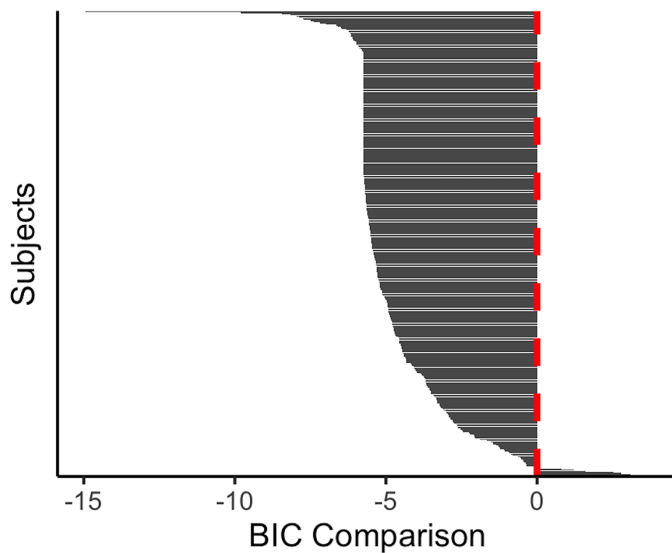
## Modeling Results

For each subject, we fit their responses with the two models using Maximum Likelihood Estimation, then evaluated which model provided a better fit to each subject's data using the Bayesian Information Criterion (BIC; Neath & Cavanaugh, 2012). For 398 of the 410 subjects, the BIC value yielded by their SDT Model fit was lower than the BIC values yielded by the Give Up Model (see Figure 4). This pattern was similar regardless of stimulus presentation condition (sequential: 201/208 preferred SDT Model; simultaneous: 201/202 preferred SDT Model). Consistent with the better-than-chance performance we found on difficult ratios, the SDT model is a more parsimonious descriptor of the data for nearly all of our participants.



Figure 4

BIC Comparison of SDT vs. Give Up Model



Note. BIC comparison < 0 indicates SDT model preferred over Give Up model for that subject.

Additionally, we also fit all 410 subjects' data together to compare the two model fits (see Figure 3 for the group-level model curves). When fitting the group data using the SDT Model, we found the best fit with  $w = .157$ , which is consistent with other published values in the literature (e.g., Halberda & Feigenson, 2008), and  $g = .214$ . When we fit the group data with the Give Up model, we found  $w = .147$ ,  $g = .231$ , and *Guess Boundary* = .004; this would indicate that subjects are expected to start guessing on a comparison of 251:250 dots. Notably, the Give Up model fit results in a curve that almost completely overlaps with the SDT model (see Figure 3). Thus, even when given the opportunity to implement a Guess Boundary at any ratio, the model essentially opts not to. A comparison of the BIC values for the SDT Model (BIC = 162,482) versus the Give Up model (BIC = 162,486.1) confirmed that the SDT Model was the most parsimonious description of the data. The small difference in BIC values between the SDT and Give Up models indicates that they obtain nearly identical fits (as expected as the Guess Boundary parameter approaches 0), but the SDT model is preferred because it has fewer parameters than the Give Up model.

## Discussion

The idea that some differences are too small to perceive has intuitive appeal. However, the data presented here suggest that humans are capable of far finer distinctions than this idea would imply. Although comparisons as large as an 8:7 ratio have been cited as the limit of approximate number perception for adults (e.g., Carey, 2009), we found that people are capable of making distinctions as fine as 50 vs. 51 dots at an above chance rate.

Although this success may feel counterintuitive, it is in fact consistent with modern models of psychophysics. For example, if representations of number are well-ordered, then there will always be some region of activation for 51 that is greater in magnitude than the activation for 50 (likewise for 101:100, etc.). If the number representations are close (e.g., 51:50), just like any two similar signals in the mind, the region of non-overlap, where 51 has a higher signal than 50, will be small, but it will be represented. And it is this small region of non-overlap in representation that drives the success we observed here – a small but significant improvement from chance. Nothing changes as one progresses to more and more difficult numerical comparisons. The observer is thinking the same thought (e.g., *who has more*), gathering the same evidence (e.g., *how many*) and making the same comparison. The only thing that *does* change is the observer's epistemic limitations for making their decision (Halberda, 2016).



In Figure 3, performance at *each and every ratio* can be described by a single parameter – the Weber fraction ( $w$ ) – which determines the amount of overlap between *any* set of numerosity representations. Recall that this argument applies whether subjects were relying solely on a numerical representation, on area, convex hull or any combination of magnitude representations.

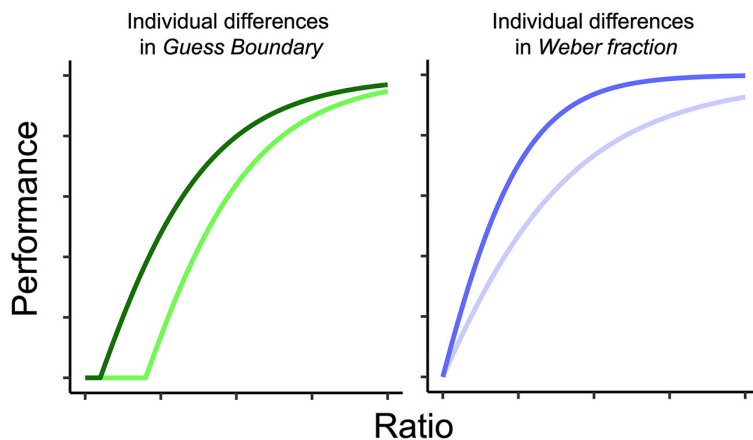
Why do some previous studies report data consistent with “at chance” performance on difficult ratios (e.g., Hauser et al., 2003; Xu & Spelke, 2000)? As mentioned in the introduction, we hypothesize that one reason may be that these studies did not have sufficient data to find support for what would undoubtedly be small effects. Here, data from 410 subjects who completed 56 trials each at the most difficult ratios gave us ample evidence to distinguish 51.3% performance from chance.

What do the present results mean for our understanding of individual differences in ANS precision? It has been typical in numerical cognition to use the point at which performance transitions from above-chance to at-chance as a metric for distinguishing between populations and species. But although such a description in terms of the minimum ratio of discriminability is widespread across the numerical literature (see Mehlis et al., 2015), we argue that the data is better represented and explained by differences in *slope*, rather than differences in *limits*. Luckily, psychophysical theories provide us with an easy way to quantify the slope of a discrimination function.

In the SDT model, individual differences are explained via the Weber fraction parameter. In such a model, while everyone would be near perfect with the easiest ratios (e.g., 2:1), performance would only drop to *chance* upon reaching equality (i.e., a ratio of 1:1), regardless of differences in precision. Some individuals may be more precise and show more rapid improvement with ratio (e.g., the darker blue curve in Figure 5, right) while others may lag behind – yet remain above chance (e.g., the lighter blue curve in Figure 5, right). This behavior looks markedly different from one defined by *limits* to what differences one can perceive (e.g., the lighter versus darker green curves in Figure 5, left). Although it may be tempting to simply categorize magnitude discrimination behavior into success or failure, the truth is in fact much more continuous.

Figure 5

Comparison of Expected Patterns of Individual Differences Based on Give Up vs. SDT Models



**Funding:** Data collection and analysis were supported by an NSF GRFP, DGE1746891, awarded to E.M.S. and a McDonnell Foundation Scholar Award awarded to J.H.

**Acknowledgments:** The authors have no additional (i.e., non-financial) support to report.

**Competing Interests:** The authors have declared that no competing interests exist.

## References

- Agrillo, C., Dadda, M., & Bisazza, A. (2007). Quantity discrimination in female mosquitofish. *Animal Cognition*, *10*(1), 63–70. <https://doi.org/10.1007/s10071-006-0036-5>
- Carey, S. (2009). *The origin of concepts*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195367638.001.0001>
- Dehaene, S. (2003). The neural basis of the Weber-Fechner law: A logarithmic mental number line. *Trends in Cognitive Sciences*, *7*(4), 145–147. [https://doi.org/10.1016/S1364-6613\(03\)00055-X](https://doi.org/10.1016/S1364-6613(03)00055-X)
- De Smedt, B., Noël, M.-P., Gilmore, C., & Ansari, D. (2013). How do symbolic and non-symbolic numerical magnitude processing skills relate to individual differences in children’s mathematical skills? A review of evidence from brain and behavior. *Trends in Neuroscience and Education*, *2*(2), 48–55. <https://doi.org/10.1016/j.tine.2013.06.001>
- Fechner, G. (1966). *Elements of psychophysics*. Holt, Rinehart & Winston. (Original work published 1860) <https://doi.org/10.1037/11304-026>
- Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences*, *8*(7), 307–314. <https://doi.org/10.1016/j.tics.2004.05.002>
- Germine, L., Nakayama, K., Loken, E., Duchaine, B., Chabris, C., Chatterjee, G., & Wilmer, J. (2010). Downloadable science: Comparing data from internet and lab-based psychology experiments. *Journal of Vision*, *10*(7), 682. <https://doi.org/10.1167/10.7.682>
- Gómez-Laplaza, L. M., & Gerlai, R. (2011). Can angelfish (*Pterophyllum scalare*) count? Discrimination between different shoal sizes follows Weber’s law. *Animal Cognition*, *14*(1), 1–9. <https://doi.org/10.1007/s10071-010-0337-6>
- Halberda, J. (2016). Epistemic limitations and precise estimates in analog magnitude representation. In A. Baron & D. Barner (Eds.), *Core knowledge and conceptual change* (pp. 167–186). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780190467630.003.0010>
- 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. <https://doi.org/10.1037/a0012682>
- Halberda, J., Ly, R., Wilmer, J. B., Naiman, D. Q., & Germine, L. (2012). Number sense across the lifespan as revealed by a massive Internet-based sample. *PNAS*, *109*(28), 11116–11120. <https://doi.org/10.1073/pnas.1200196109>
- Halberda, J., Mazocco, M. M. M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature*, *455*(7213), 665–668. <https://doi.org/10.1038/nature07246>
- Hauser, M. D., Tsao, F., Garcia, P., & Spelke, E. S. (2003). Evolutionary foundations of number: Spontaneous representation of numerical magnitudes by cotton-top tamarins. *Proceedings of the Royal Society of London: Series B. Biological Sciences*, *270*(1523), 1441–1446. <https://doi.org/10.1098/rspb.2003.2414>
- Izard, V., Sann, C., Spelke, E. S., & Streri, A. (2009). Newborn infants perceive abstract numbers. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(25), 10382–10385. <https://doi.org/10.1073/pnas.0812142106>
- Laming, D. R. J. (1973). *Mathematical psychology*. Academic Press.
- Mazocco, M. M. M., Feigenson, L., & Halberda, J. (2011). Preschoolers’ precision of the approximate number system predicts later school mathematics performance. *PLoS ONE*, *6*(9), Article e23749. <https://doi.org/10.1371/journal.pone.0023749>
- McGill, W. J., & Goldberg, J. P. (1968). A study of the near-miss involving Weber’s law and pure-tone intensity discrimination. *Perception & Psychophysics*, *4*(2), 105–109. <https://doi.org/10.3758/BF03209518>
- Mehlis, M., Thünken, T., Bakker, T. C. M., & Frommen, J. G. (2015). Quantification acuity in spontaneous shoaling decisions of three-spined sticklebacks. *Animal Cognition*, *18*(5), 1125–1131. <https://doi.org/10.1007/s10071-015-0884-y>
- Neath, A. A., & Cavanaugh, J. E. (2012). The Bayesian information criterion: Background, derivation, and applications. *Wiley Interdisciplinary Reviews: Computational Statistics*, *4*(2), 199–203. <https://doi.org/10.1002/wics.199>
- Norris, J. E., & Castronovo, J. (2016). Dot display affects approximate number system acuity and relationships with mathematical achievement and inhibitory control. *PLoS ONE*, *11*(5), Article e0155543. <https://doi.org/10.1371/journal.pone.0155543>
- Odic, D., Hock, H., & Halberda, J. (2014). Hysteresis affects approximate number discrimination in young children. *Journal of Experimental Psychology: General*, *143*(1), 255–265. <https://doi.org/10.1037/a0030825>
- Park, J., & Brannon, E. M. (2013). Training the approximate number system improves math proficiency. *Psychological Science*, *24*(10), 2013–2019. <https://doi.org/10.1177/0956797613482944>

- Piazza, M., Facoetti, A., Trussardi, A. N., Berteletti, I., Conte, S., Lucangeli, D., Dehaene, S., & Zorzi, M. (2010). Developmental trajectory of number acuity reveals a severe impairment in developmental dyscalculia. *Cognition*, *116*(1), 33–41. <https://doi.org/10.1016/j.cognition.2010.03.012>
- Pica, P., Lemer, C., Izard, V., & Dehaene, S. (2004). Exact and approximate arithmetic in an Amazonian indigene group. *Science*, *306*(5695), 499–503. <https://doi.org/10.1126/science.1102085>
- Qiu, K., Chen, E. H., Wan, S., & Bailey, D. H. (2021). Multilevel meta-analysis on the causal effect of approximate number system training on symbolic math performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *47*(11), 1820–1835. <https://doi.org/10.1037/xlm0001087>
- Starr, A., DeWind, N. K., & Brannon, E. M. (2017). The contributions of numerical acuity and non-numerical stimulus features to the development of the number sense and symbolic math achievement. *Cognition*, *168*, 222–233. <https://doi.org/10.1016/j.cognition.2017.07.004>
- Szűcs, D., & Myers, T. (2017). A critical analysis of design, facts, bias and inference in the approximate number system training literature: A systematic review. *Trends in Neuroscience and Education*, *6*(August 2016), 187–203. <https://doi.org/10.1016/j.tine.2016.11.002>
- 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. <https://doi.org/10.1007/s10071-003-0167-x>
- Wang, J., Halberda, J., & Feigenson, L. (2017). Approximate number sense correlates with math performance in gifted adolescents. *Acta Psychologica*, *176*(April), 78–84. <https://doi.org/10.1016/j.actpsy.2017.03.014>
- Wang, J., Odic, D., Halberda, J., & Feigenson, L. (2016). Changing the precision of preschoolers' approximate number system representations changes their symbolic math performance. *Journal of Experimental Child Psychology*, *147*, 82–99. <https://doi.org/10.1016/j.jecp.2016.03.002>
- Weber E. H. (1978). *The sense of touch (De tactu*, transl. by Ross H. E.; *Der Tastsinn*, transl. by Murray D. J.). Academic Press. (Original work published 1834)
- Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition*, *74*(1), B1–B11. [https://doi.org/10.1016/S0010-0277\(99\)00066-9](https://doi.org/10.1016/S0010-0277(99)00066-9)



*Journal of Numerical Cognition* (JNC) is an official journal of the Mathematical Cognition and Learning Society (MCLS).



leibniz-psychology.org

PsychOpen GOLD is a publishing service by Leibniz Institute for Psychology (ZPID), Germany.