



Numerical distance effect size is a poor metric of approximate number system acuity

Dana Chesney¹

Published online: 12 April 2018
© The Psychonomic Society, Inc. 2018

Abstract

Individual differences in the ability to compare and evaluate nonsymbolic numerical magnitudes—approximate number system (ANS) acuity—are emerging as an important predictor in many research areas. Unfortunately, recent empirical studies have called into question whether a historically common ANS-acuity metric—the size of the numerical distance effect (NDE size)—is an effective measure of ANS acuity. NDE size has been shown to frequently yield divergent results from other ANS-acuity metrics. Given these concerns and the measure’s past popularity, it behooves us to question whether the use of NDE size as an ANS-acuity metric is theoretically supported. This study seeks to address this gap in the literature by using modeling to test the basic assumption underpinning use of NDE size as an ANS-acuity metric: that larger NDE size indicates poorer ANS acuity. This assumption did not hold up under test. Results demonstrate that the theoretically ideal relationship between NDE size and ANS acuity is not linear, but rather resembles an inverted J-shaped distribution, with the inflection points varying based on precise NDE task methodology. Thus, depending on specific methodology and the distribution of ANS acuity in the tested population, positive, negative, or null correlations between NDE size and ANS acuity could be predicted. Moreover, peak NDE sizes would be found for near-average ANS acuities on common NDE tasks. This indicates that NDE size has limited and inconsistent utility as an ANS-acuity metric. Past results should be interpreted on a case-by-case basis, considering both specifics of the NDE task and expected ANS acuity of the sampled population.

Keywords Numerical distance effect · Estimation · Approximate number system · Acuity · Numerical magnitudes

People can evaluate nonsymbolic numerical magnitudes (i.e., which pack has more wolves) without counting (Taves, 1941) via what is known as the approximate number system (ANS). The ANS allows us to perceive numerical magnitudes from the world in an analog fashion, similarly to how we perceive other magnitudes, like size (Kaufman, Lord, Reese, & Volkman, 1949). This skill, “ANS acuity,” varies: Some individuals can make faster and more accurate judgments than others (Halberda & Feigenson, 2008). Better ANS acuity has been linked to better math skills and better standardized test performance (Gilmore, McCarthy, & Spelke, 2010; Halberda,

Mazzocco, & Feigenson, 2008) (also see Chen & Li, 2014, for meta-analysis and review) and may even influence judgment and decision-making in adults (Peters, Slovic, Västfjäll, & Mertz, 2008; Schley & Peters, 2014). Unfortunately, recent empirical studies call into question the effectiveness of a historically common ANS-acuity metric: the size of the numerical distance effect (NDE size; Gilmore, Attridge, & Inglis, 2011; Holloway & Ansari, 2009; Inglis & Gilmore, 2014; Maloney, Risko, Preston, Ansari, & Fugelsang, 2010; also see Sasanguie, Defever, Van den Bussche, & Reynvoet, 2011). The goal of this study was to assess the theoretical support for using NDE size as an ANS-acuity metric.

Dana Chesney was at The Ohio State University and is now at St. John’s University. This research was supported by NSF Grant SES-1155924. Thank you to Ellen Peters for her help and support. Note: An earlier version of this paper was presented at CogSci 2016.

✉ Dana Chesney
chesneyd@stjohns.edu

Individual differences, ANS acuity, and the NDE

There is strong evidence that people invoke ANS-based analog magnitudes when considering symbolic numbers (Dehaene, 1992; Dehaene, Bossini, & Pascal, 1993; Moyer & Landauer,

¹ Department of Psychology, St. John’s University, 8000 Utopia Pkwy, Jamaica, NY 11439, USA

1967). Moyer and Landauer's (1967) seminal study demonstrated that people show distance effects when making quantity judgments about symbolic magnitudes. For example, people are faster at determining that 6 is smaller than 9 than they are at determining that 7 is smaller than 8. Such effects are a classic pattern in analog magnitude comparisons (Moyer & Landauer, 1967).

Moyer and Landauer's (1967) now classic approach of using the presence of distance effects to demonstrate that the ANS is invoked in symbolic magnitude comparisons appears to have inspired the later use of NDE size as an ANS-acuity metric. In a practice that seems to originate with Sekuler and Mierkiewicz (1977), researchers will calculate NDE size by finding the savings in the speed and/or accuracy of numerical comparisons (e.g., "Which is larger?") at larger (easier) versus smaller (harder) distances. Larger NDE size is taken to indicate poorer ANS acuity (Peters et al., 2008). Recently, several studies have questioned this measure's ability to distinguish individual differences in ANS acuity (Gilmore et al., 2011; Holloway & Ansari, 2009; Inglis & Gilmore, 2014; Maloney et al., 2010; also see Sasanguie et al., 2011) (also see Price, Palmre, Battista, & Ansari, 2012). Given these issues of empirical support, this manuscript seeks to address whether the use of NDE size as an ANS-acuity metric is theoretically supported.

ANS theory and NDE size

The exact nature of the ANS has yet to be completely determined, but it is well established that it obeys Weber's law (Cordes, Gelman, Gallistel, & Whalen, 2001; Dehaene, Izard, Spelke, & Pica, 2008; Mechner, 1958; Meck & Church, 1983; Whalen, Gallistel, & Gelman, 1999). As is typically the case for magnitude perception (see Kingdom & Prins, 2010), numerical magnitudes are not perceived exactly. Rather, percepts are normally or quasinormally distributed around a (possibly biased) mean value. The ability to distinguish between two quantities is dependent on the amount of overlap between their perceived magnitude distributions. Importantly, this overlap—and thus the ease with which two values can be distinguished—is dependent upon their ratio. As a result, one can observe both distance and size effects in magnitude discriminations. It is easier to distinguish numerical quantities that are more *distant* from each other (6 dots [:::] vs. 12 dots [:::::]) than those that are closer together (8 dots [::::] vs. 10 dots [:::::]). Also, it is easier to distinguish numerical quantities at the same distance with smaller *sized* magnitudes (6 dots [:::] vs. 8 dots [::::]) than with larger magnitudes (14 dots [:::::] vs. 16 dots [:::::]). ANS magnitude comparisons yield standard psychophysical functions: The likelihood that an individual will successfully discriminate between two magnitudes will increase curvilinearly from chance to asymptote at or near 100% accuracy, as the ratio of the larger to the smaller value increases.

Reaction Times (RTs) similarly decrease with the comparison ratio (Whalen et al., 1999; see Kingdom & Prins, 2010, for a discussion of psychophysical functions).

ANS acuity is defined by an individual's Weber fraction (Cordes et al., 2001; Dehaene, et al., 2008; Halberda et al., 2008; Siegler & Opfer, 2003; Whalen et al., 1999). Weber's law implies that the standard deviation (*SD*) of the distribution around an estimated magnitude is a constant proportion of that magnitude's mean (*M*). This constant proportion is, by definition, the Weber fraction (*w*) of the perceiver's ANS. This results in greater overlap between the magnitude distributions perceived from stimuli at smaller ratios (:::/:::, 1.33) than at larger ratios (:::/::::, 2). After accounting for other biases, *w* determines the variability in the representation of a particular magnitude, which in turn determines the amount of overlap between any two magnitudes, which finally determines how likely it is that an individual will be able to tell two nonsymbolic magnitudes apart (ANS acuity). The smaller an individual's *w*, the better the individual will be at discriminating between nonsymbolic numerical magnitudes because there is less overlap in their numerical magnitude perceptions.

It follows that the ANS's contribution to NDE size should be a function of the specific magnitudes being compared and ANS acuity (*w*). Thus, we can derive the relative size of the ANS's contribution to NDE size for any specific task and any given *w* by considering the resulting distributions. As long as judgments are based on ANS distributions, error rates and RTs should be functionally related to the amount of overlap in these distributions.

Method

The goal of this work is to assess the theoretical support for using NDE size as an ANS acuity metric. Thus, the theoretically ideal NDE sizes calculated here are dependent only on the magnitude ratios of the stimuli and *w*. Real-world data would involve other sources of RT and error (attention to task, nondecision time, etc.), adding noise that would make this relationship less clear. However, as these factors are separate from the ANS, they are excluded from this ideal model.

Formula: The relationship of overlap and *erfc* with *w*

Calculations are based on the linear model of the ANS, which claims that means of perceived numerical magnitude distributions increase linearly with the size of the stimuli, and the distributions' standard deviations are proportional to their means (i.e., scalar variability; Cordes et al., 2001). (Note: Magnitudes might alternatively be modeled as logarithmically spaced with constant variability, yielding similar outcomes.) Thus, the ratio of the standard deviation to the mean is constant for a given individual on a given task. This constant is the

w of the individual's ANS: their ANS acuity. Here, magnitudes are treated as Gaussian distributions around unbiased means equal to the stimulus value (M), where $SD = w \times M$. Thus, the derived overlap in ANS distributions is a function of the stimulus ratio and the Weber fraction (w) of the ANS. Overlap calculations are described in the [Appendix](#). Additionally, following the method used by Halberda et al. (2008), the *erfc* (the complementary error function) was used to determine the rate at which a given pair of magnitudes will not be distinguished. Given no other sources of error, *erfc* should be equal to twice the error rate of ANS-based magnitude judgments, as the observer would be presumed to choose the correct answer by chance on half of such trials. The MATLAB code used for these calculations is given in the [Appendix](#).

Overlaps and *erfcs* can be calculated for any stimulus ratio and w . Thus, the theoretically maximum contribution of ANS acuity to NDE size can be found for any w on any particular NDE task. However, the ranges of greatest interest are those that correspond to w s seen in humans. Consistent with prior work (e.g., Cordes et al., 2001; Whalen et al., 1999), Chesney, Bjälkebring, & Peters, (2015) found mean w s of .22 ($SD = .06$). However, others have found mean w s of .11 in educated adults (Dehaene et al., 2008). ANS acuity also varies with age. Halberda et al. (2008) found that 14-year-olds had mean w s of .28 ($SD = .10$). Studies with infants have found w s of 1.0 (Xu & Spelke, 2000), while 1-year-olds show w s of less than .5 (Cantrell & Smith, 2013).

Task, NDE size, and ANS acuity

The ideal relationship between NDE size and w was modeled for two different tasks and calculation methods.

Task 1 Task 1 is based on an NDE measure like that used by Peters et al. (2008). Participants are given a central comparison value (e.g., 5) and asked to indicate whether stimulus values are greater or less than that value. The stimuli follow a 2×2 design: Half are less (e.g., 1, 4) and half are greater (e.g., 6, 9) than the central value. Also, half are close (e.g., 4, 6) and half are far (e.g., 1, 9) from the central value. An individual's NDE size is operationalized as the difference in accuracy and/or RT on close versus far trials.

Here, NDE size for a given w and pair of stimulus ratios was calculated as the difference in the modeled overlaps and *erfcs*:

$$\begin{aligned} \text{Overlap difference} &: \text{Overlap}(\text{small ratio}) - \text{Overlap}(\text{large ratio}) \\ \text{erfc difference} &: \text{erfc}(\text{small ratio}) - \text{erfc}(\text{large ratio}). \end{aligned}$$

In this paradigm, although stimuli distances are symmetrical around the central comparison value, the ratios are asymmetrical. For the stimuli greater than the central

value (high), the close and far ratios are 6/5 (1.2) and 9/5 (1.8), respectively. For the stimuli less than the central value (low), the ratios are 5/4 (1.25) and 5/1 (5). Analyses used in the literature (e.g., Peters et al., 2008) classify stimuli as near and far, collapsing across these ratio differences. This is modeled here by averaging NDE sizes found for ratios above and below the central value (average).

Task 2 An alternative method of gauging NDE size is to find the slope of the linear regression of RTs or error rates on ratio or distance, treating ratio/distance as continuous rather than dichotomous (e.g., Sekuler & Mierkiewicz, 1977). Negative slopes indicate the presence of a distance effect. Larger (i.e., more strongly negative) absolute slopes are treated as indicating larger w s (i.e., poorer ANS acuity). Theoretically, ideal NDE slopes were modeled based on the comparison task developed by Chesney et al. (2015), which used ratios between 1 and 2.6 and numerical magnitudes between 10 and 30. While magnitude overlaps are dependent on ratio, NDE slope calculations have used distance as the Independent variable (e.g., Sekuler & Mierkiewicz, 1977). Therefore NDE slopes were found by regressing overlap and *erfc* on both the ratio and absolute distance between comparison pairs across the human range of w s.

Results

Modeled overlaps and *erfcs*

Figure 1 illustrates the modeled ideal ANS magnitude overlaps and *erfcs* for w s from .04 to .48 (w ranges seen in typically developing adults and older children) and ratios from 1 to 5 (greater/lesser), covering the range of difficulty from *impossible* to *easy* across these w s. Smaller ratios have greater overlaps and *erfcs* than larger ratios. As ratio increases, there is a steep initial drop for both overlaps and *erfcs*, which then “turns the corner” to asymptote to zero. All w s yield this same pattern, but the initial drop is steeper and the “corner” reached faster for smaller w s.

The relationship of overlap and *erfc* to w predicts a nonlinear relationship between NDE size and w . This is shown in Fig. 2. People with w s of .12 would have a lot of savings discriminating 9 versus 12 compared with 9 versus 10, as they would find 9 versus 12 easy, but 9 versus 10 would be at the upper range of their skill. However, people with w s of .32 would have *less* savings, as they would find both 9 versus 12 and 9 versus 10 to be difficult. People with w s of .04 would have even less savings because they would find both 9 versus 12 and 9 versus 10 to be very easy.

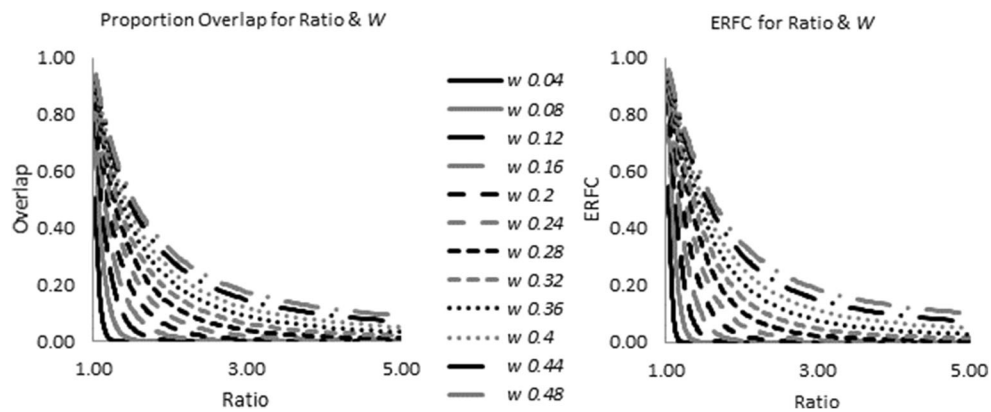


Fig. 1 Derived ANS magnitude overlaps (left) and *erfcs* (right) for w s ranging from .04 (excellent acuity) to .48 (poorer acuity), at high/low comparison value ratios ranging from 1 (equal values) to 5 (e.g., 50 vs. 10)

Relationship between NDE size and w is J shaped—

Task 1 Modeled NDE sizes for the expected human range of w s (i.e., from skilled adults' near 0, to infants' 1) on Task 1 are presented in Fig. 3. As can be seen, the ideal relationship between w and NDE size is not linear, but follows an inverted J-shaped curve. A strong positive linear relationship between w and NDE size only exists for w s ranging between $\sim .05$ and $\sim .20$, quickly rising from near zero savings to a 40% overlap savings, and 45% *erfc* savings at w s of .20. Both overlap and

erfc savings then are near flat between w s of $\sim .20$ and $\sim .60$, with overlap savings peaking at 49% for w s of .39 and *erfc* savings peaking at 54% for w s of .34. Past these peaks, savings decline slowly as w increases.

Clearly, the general presumption that larger NDE sizes correlate with larger w s does not always hold. One could expect NDE size and w to have a positive linear relationship only if the population's w s were located between .05 and .20. Indeed, depending on the population's w distribution, one could predict a positive, negative, or nonexistent correlation between

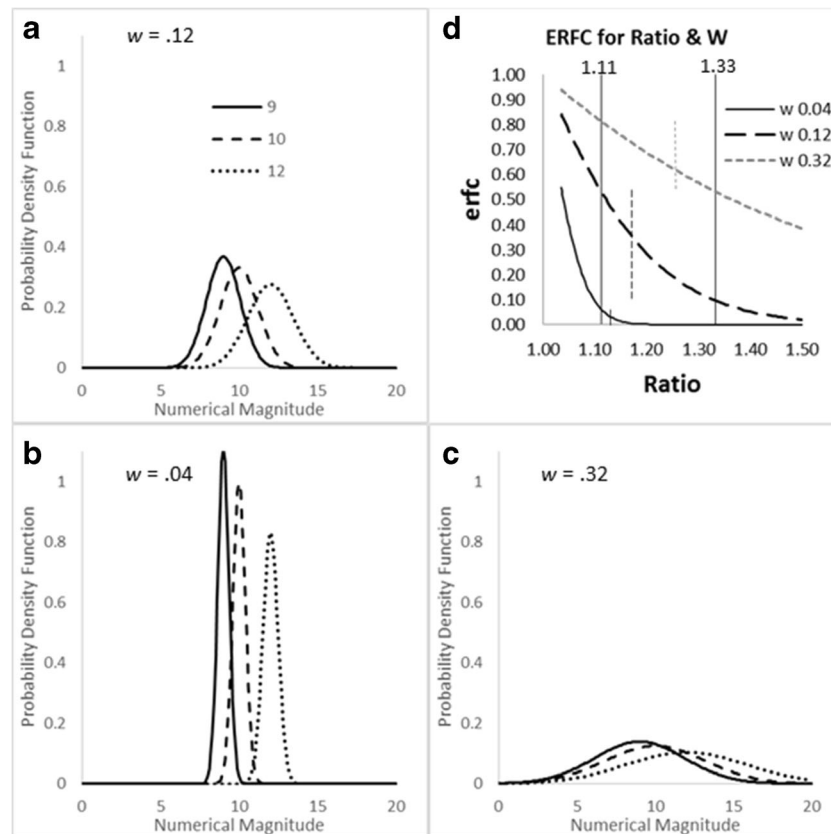


Fig. 2 a–c Magnitude distributions of 9, 10, and 12 at w s .12, .04, and .32. **d** *Erfcs* for a range of ratios including 10/9 (1.11) and 12/9 (1.33) for these w s. Vertical segments illustrate the size of the *erfc* differences between these ratios for the w s

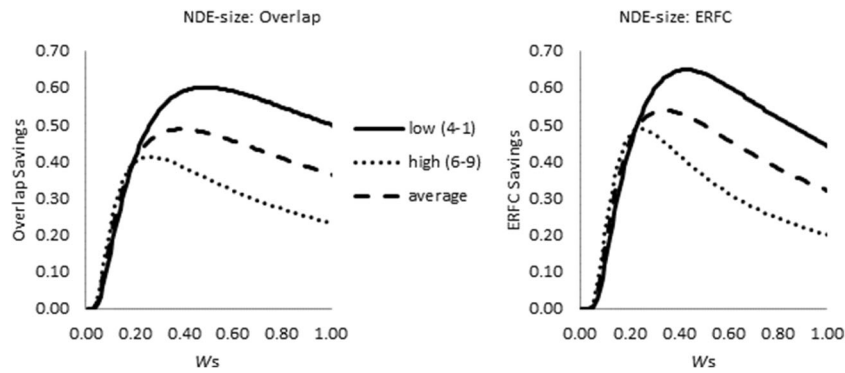


Fig. 3 Derived NDE sizes for overlaps (left) and *erfcs* (right) on Task 1

NDE size and *w*. Moreover, one could not necessarily recover *ws* from NDE size because, again owing to the J-shaped relation, more than one *w* maps to the same NDE size. For example, average *erfc* savings of .48 map to *ws* of both .22 (average adult) and .55 (very poor).

Task 2 The J-shaped relationship is not task dependent. Ratio regressions for Task 2 yielded a stronger relationship between NDE slope and *w* than absolute distance, but both yielded a J-shaped curve (see Fig. 4). A strong linear relationship between *w* and NDE slope only held for *ws* up to ~.2, with an inflection point at *ws* of .32. These curves are very similar to those calculated for a task comparing all possible unequal pairings of the values 1–9.

Discussion

Even though numerical distance effects can indicate the involvement of the ANS in a task, NDE tasks have limited utility for measuring individual differences in ANS acuity. This model provides a novel theoretical exploration of why this is the case: one cannot a priori expect NDE size and ANS acuity to be linearly related. The J-shaped relationship between *w* and NDE size persists across tasks and analytical methods, although the inflection points are task specific. Small NDE sizes are expected both for individuals with

particularly small and particularly large *ws*. As a result, even the direction of the correlation would be dependent both on the specifics of the task and on the *w* distribution in the sample.

For typical NDE tasks, like those above, peak NDE size is approached at *ws* of ~.2–.3. The location of this peak is a real concern, as several studies have found adults’ *ws* typically center around ~.22 (e.g., Cordes et al., 2001; Whalen et al., 1999). But the range of human *ws* is wide. Other studies have found mean *ws* of .11 in adults (Dehaene et al., 2008) and *ws* of 1.0 in infants (Xu & Spelke, 2000). Thus, one cannot presume the *w* range in a novel population will coincide with the *w* range in which the relationship of *w* and NDE size is quasilinear. This is problematic to the literature as a whole, and particularly for research attempting to draw conclusions about the nature of ANS acuity’s involvement in other cognitive tasks.

Nevertheless, if the *w* range within a to-be-tested population is known, one might construct a task for which the assumption of a near-linear relationship between NDE size and ANS acuity is theoretically supported. One could expect *w* ranges of .10 to .34 in American university students (Chesney et al., 2015; but see Dehaene et al., 2008). Over this range, NDE size found using the ratios 5 versus 1.25 would yield a strong linear relationship between ideal NDE size and *w* (this is illustrated by the “low” line on Fig. 3). Stimuli should be carefully chosen to avoid confounds. For example,

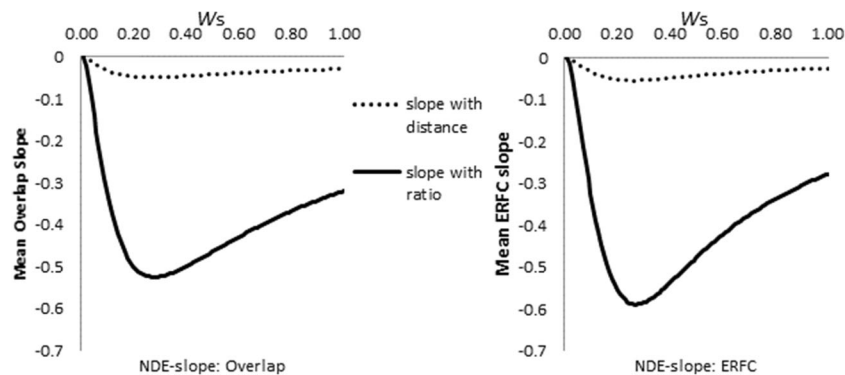


Fig. 4 Derived NDE slopes for overlaps (left) and *erfcs* (right) on Task 2

nonsymbolic stimuli (e.g., dot sets) should be used: Symbolic number knowledge could interfere with task performance, and, indeed, there is some debate as to whether distance effects seen with symbolic number are necessarily the result of ANS involvement (Krajcsi, Lengyel, & Kojouharova, 2016). Nonnumeric properties of the stimuli (e.g., area) should be carefully controlled, as these are known to influence ANS assessments (Hurewitz, Gelman, & Schnitzer, 2006). Further, ratios should be instantiated using quantities sufficiently large to avoid interference from subitizing (a process that allows individuals to assess small quantities—typically less than 7—accurately without counting; see Chesney & Haladjian, 2011; Feigenson, Dehaene, & Spelke, 2004). However, such tasks should still be tested empirically. Other factors, such as individual differences in nondecision time and error tolerance, might mask the theoretical quasilinear relation in practice. A better course might be to assess ANS acuity using tested tasks that find w by fitting numerosity comparison performance to sigmoidal or psychophysical curves (e.g., Chesney et al., 2015; Halberda et al., 2008).

Conclusions

It is important that ANS-acuity metrics are both empirically and theoretically supported. As shown here, the assumed linear relationship between NDE size and ANS acuity is only theoretically supported in some conditions, and researchers may not be able to discern whether these conditions have been met. Considered in combination with the questionable empirical support of NDE size as an ANS-acuity metric (Gilmore et al., 2011; Holloway & Ansari, 2009; Inglis & Gilmore, 2014; Maloney et al., 2010; also see Sasanguie et al., 2011), it is recommended that use of this metric should be avoided. Interpretation of existing NDE-size data should take into account the expected relationship between NDE size and ANS acuity for that specific methodology and typical w ranges found for age-matched and education-matched populations.

Appendix

Overlap calculations are based on finding the value at which it is equally likely that a magnitude from the lower distribution would be greater than that value, and that a magnitude from the higher distribution would be less than that value. This value becomes a cut point separating the “body” and “tails” of these magnitude distributions. The overlap is defined as the proportion of distributions in these equal tails divided by the remainder of the distributions.

MATLAB code used to calculate the proportion of overlap in the ANS distributions of any two values, given w :

$$\text{invCdf} = 1 - \text{normcdf}((\text{High} - \text{Low}) / (w \times (\text{High} + \text{Low})), 0, 1)$$

$$\text{overlap} = (2 \times \text{invCdf}) / (2 - (2 \times \text{invCdf})).$$

The *erfc* calculation followed the process given in Halberda et al. (2008). MATLAB code used to find the value of the complimentary error function (*erfc*: the rate at which the ANS values will not be distinguished, double the ideal error rate):

$$\text{erfc}(\text{abs}(\text{High} - \text{Low}) / (\text{sqrt}((\text{High}^2) + (\text{Low}^2)) \times \text{sqrt}(2) \times w))$$

Variables “high” and “low” referred to the higher and lower value in a comparison pair, respectively (e.g., 6 & 5). Variable “w” referred to the Weber fraction (i.e., ANS acuity).

The overlap calculation found the point at which the lower tail of the higher value’s distribution was equal to the higher tail of the lower value’s distribution, and then calculated the sum of the tails divided by the sum of the bodies.

References

- Cantrell, L. M., & Smith, L. B. (2013). Set size, individuation, and attention to shape. *Cognition*, *126*, 258–267. <https://doi.org/10.1016/j.cognition.2012.10.007>
- Chen, Q., & Li, J. (2014). Association between individual differences in non-symbolic number acuity and math performance: A meta-analysis. *Acta Psychologica*, *148*, 163–172. <https://doi.org/10.1016/j.actpsy.2014.01.016>
- Chesney, D. L., Bjälkebring, P., & Peters, E. (2015). How to estimate how well people estimate: Evaluating measures of individual differences in the approximate number system. *Attention, Perception, & Psychophysics*, *77*, 2781–2802. <https://doi.org/10.3758/s13414-015-0974-6>
- Chesney, D. L., & Haladjian, H. H. (2011). Evidence for a shared mechanism used in multiple-object tracking and subitizing. *Attention, Perception, & Psychophysics*, *73*, 2457–2480. <https://doi.org/10.3758/s13414-011-0204-9>
- Cordes, S., Gelman, R., Gallistel, C. R., & Whalen, J. (2001). Variability signatures distinguish verbal from nonverbal counting for both large and small numbers. *Psychonomic Bulletin & Review*, *8*, 698–707. <https://doi.org/10.3758/BF03196206>
- Dehaene, S. (1992). Varieties of numerical abilities. *Cognition*, *44*, 1–42. [https://doi.org/10.1016/0010-0277\(92\)90049-N](https://doi.org/10.1016/0010-0277(92)90049-N)
- Dehaene, S., Bossini, S., & Pascal, G. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General*, *122*, 371–396. <https://doi.org/10.1037/0096-3445.122.3.371>
- Dehaene, S., Izard, V., Spelke, E., & Pica, P. (2008). Log or linear? Distinct intuitions of the number scale in Western and Amazonian indigene cultures. *Science*, *320*, 1217–1220. <https://doi.org/10.1126/science.1156540>
- Feigenson, L., Dehaene, S., & Spelke, E. S. (2004). Core systems of number. *Trends in Cognitive Sciences*, *8*, 307–314. <https://doi.org/10.1016/j.tics.2004.05.002>
- Gilmore, C., Attridge, N., & Inglis, M. (2011). Measuring the approximate number system. *The Quarterly Journal of Experimental*

- Psychology*, 64, 2009–2109. <https://doi.org/10.1080/17470218.2011.574710>
- Gilmore, C.K., McCarthy, S.E., & Spelke, E.S. (2010). Non-symbolic arithmetic abilities and mathematics achievement in the first year of formal schooling. *Cognition*, 115, 394–406. <https://doi.org/10.1016/j.cognition.2010.02.002>
- Halberda, J., & Feigenson, L. (2008). Developmental change in the acuity of the “number sense”: The approximate number system in 3-, 4-, 5-, 6-year-olds and adults. *Developmental Psychology*, 44, 1457–1465. <https://doi.org/10.1037/a0012682>
- Halberda, J., Mazocco, M., & Feigenson, L. (2008). Individual differences in nonverbal number acuity predict maths achievement. *Nature*, 455, 665–668. <https://doi.org/10.1038/nature07246>
- Holloway, I. D., & Ansari, D. (2009). Mapping numerical magnitudes onto symbols: The numerical distance effect and individual differences in children’s mathematics achievement. *Journal of Experimental Child Psychology*, 103, 17–29. <https://doi.org/10.1016/j.jecp.2008.04.001>
- Hurewitz, F., Gelman, R., & Schnitzer, B. (2006). Sometimes area counts more than number. *Proceedings of the National Academy of Sciences*, 103, 19599–19604. <https://doi.org/10.1073/pnas.0609485103>
- Inglis, M., & Gilmore, C. (2014). Indexing the approximate number system. *Acta Psychologica*, 145, 147–155. <https://doi.org/10.1016/j.actpsy.2013.11.009>
- Kaufman, E. L., Lord, M. W., Reese, T. W., & Volkman, J. (1949). The discrimination of visual number. *American Journal of Psychology*, 62, 498–525. <https://doi.org/10.2307/1418556>
- Kingdom, F. A. A., & Prins, N. (2010). *Psychophysics: A practical introduction*. London, UK: Academic Press.
- Krajcsi, A., Lengyel, G., & Kojouharova, P. (2016). The source of the symbolic numerical distance and size effects. *Frontiers in Psychology*, 7. <https://doi.org/10.3389/fpsyg.2016.01795>
- Maloney, E. A., Risko, E. F., Preston, F., Ansari, D., & Fugelsang, J. A. (2010). Challenging the reliability and validity of cognitive measures: The case of the numerical distance effect. *Acta Psychologica* 134, 154–161. <https://doi.org/10.1016/j.actpsy.2010.01.006>
- Mechner, F. (1958). Probability relations within response sequence maintained under ratio reinforcement. *Journal of the Experimental Analysis of Behavior*, 1, 109–121. <https://doi.org/10.1901/jeab.1958.1-109>
- Meck, W. H., & Church, R. M. (1983). A mode control model of counting and timing processes. *Journal of Experimental Psychology: Animal Behavior Processes*, 9, 320–334. <https://doi.org/10.1037/0097-7403.9.3.320>
- Moyer, R. S., & Landauer, T. K. (1967). Time required for judgements of numerical inequality. *Nature*, 215, 1519–1520. <https://doi.org/10.1038/2151519a0>
- Peters, E., Slovic, P., Västfjäll, D., & Mertz, C. K. (2008). Intuitive numbers guide decisions. *Judgment and Decision Making*, 3, 619–635. Retrieved from <http://journal.sjdm.org/8827/jdm8827.pdf>
- Price, G. R., Palmer, D., Battista, S., & Ansari, D. (2012). Nonsymbolic numerical magnitude comparison: Reliability and validity of different task variants and outcome measures, and their relationship to arithmetic achievement in adults. *Acta Psychologica*, 140, 50–57. <https://doi.org/10.1016/j.actpsy.2012.02.008>
- Sasanguie, D., Defever, E., Van den Bussche, E., & Reynvoet, B. (2011). The reliability of and the relation between non-symbolic numerical distance effects in comparison, same-different judgments and priming. *Acta Psychologica*, 136, 73–80. <https://doi.org/10.1016/j.actpsy.2010.10.004>
- Schley, D. R., & Peters, E. (2014). Assessing “economic value”: Symbolic number mappings predict risky and riskless valuations. *Psychological Science*, 25, 753–761. <https://doi.org/10.1177/0956797613515485>
- Sekuler, R., & Mierkiewicz, D. (1977). Children’s Judgments of Numerical Inequality. *Child Development*, 48, 630–633. <https://doi.org/10.2307/1128664>
- Siegler, R. S., & Opfer, J. E. (2003). The development of numerical estimation: Evidence for multiple representations of numerical quantity. *Psychological Science*, 14, 237–243. <https://doi.org/10.1111/1467-9280.02438>
- Taves, E. H. (1941). Two mechanisms for the perception of visual numerosness. *Archives of Psychology*, 37(265), 1–47.
- Whalen, J., Gallistel, C. R., & Gelman, R. (1999). Non-verbal counting in humans: The psychophysics of number representation. *Psychological Science*, 10, 130–137. <https://doi.org/10.1111/1467-9280.00120>
- Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition*, 74, B1–B11. [https://doi.org/10.1016/S0010-0277\(99\)00066-9](https://doi.org/10.1016/S0010-0277(99)00066-9)