The Relationship Between the Numerical Distance Effect and Approximate Number System Acuity is Non-Linear

Dana Chesney (chesneyd@stjohns.edu)
Department of Psychology, St. John's University, 8000 Utopia Pkwy.,
Jamaica, NY 11439 USA

Abstract

People can estimate numerical quantities, like the number of grapes in a bunch, using the Approximate Number System (ANS). Individual differences in this ability (ANS acuity) are emerging as an important predictor in research areas ranging from math skills to judgment and decision making. One commonly used ANS acuity metric is the size of the Numerical Distance Effect (NDE): the amount of savings in RT or errors when distinguishing stimuli values as the numerical distance between them increases. However, the validity of this metric has recently been questioned. Here, we model the relationship between the NDE-size and ANS acuity. We demonstrate that the relationship between NDE-size and ANS acuity should not be linear, but rather should resemble an inverted J-shaped distribution, with the largest NDE-sizes typically being found for near average ANS acuities.

Keywords: Numerical Distance effect; Estimation; Approximate Number system

Introduction

People can evaluate non-symbolic numerical magnitudes (i.e., which pack has more wolves) without counting (Taves, 1941). Indeed, people can respond based on these perceived magnitudes without necessarily linking these values to symbolic numbers (Kaufman, Lord, Reese, & Volkmann, 1949). The system that makes these non-symbolic numerical magnitude evaluations is herein referred to as the Approximate Number System (ANS). Essentially, 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 et al.). The facility of the ANS to make numerical magnitude judgements (ANS 'acuity') is thought to vary among individuals, such that some individuals can make faster and more accurate judgments with smaller magnitude ratios than other individuals (Halberda & Feigenson, 2008). Better ANS acuity has been linked to better math skills in elementary schools and better performance on standardized tests (Gilmore, McCarthy, & Spelke, 2010; Halberda et al., 2008). Recent work has even suggested that the ability to understand the magnitude of numbers, a skill at least partially grounded in ANS acuity, influences judgment and decision making in adults (Peters, et al., 2008; Schley & Peters, 2014). As such, it is important that researchers use valid and reliable measures to assess ANS acuity. Unfortunately, while assessments of mathematical skill are well understood, being similar to math tests one might take at school (e.g. Cokely, Galesic, Schulz, Ghazal, & Garcia-Retamero, 2012; Lipkus, Samsa, & Rimer, 2001; Weller, Dieckmann, Tusler, Mertz, Burns & Peters, 2013), metrics of individual differences in ANS acuity are less well investigated.

One widely used metric of ANS acuity is the 'size' of the Numerical Distance Effect (NDE): The difference in reaction time or accuracy between distinguishing values that are close to each other and distinguishing values that are more distant from each other. In a practice which seems to have which originated with Sekuler and Mierkiews (1977), researchers will measure the speed and accuracy of numerical comparisons (e.g. "Which is larger?") at smaller (harder) and larger (easier) distances. Bigger differences in speed and accuracy between these easier and harder trials are interpreted as evidence of poorer ANS acuity (see Price, Palmre, Battista, & Ansari, 2012 for discussion). Recently, NDE-size has come under fire, with several studies questioning both its reliability, and its ability to distinguish individual differences in ANS acuity (Gilmore, Attride, & Inglis, 2011; Holloway & Ansari, 2009; Inglis & Gilmore, 2014; Maloney, Risko, Preston, Ansari, & Fugelsang, 2010; also see Sasanguie, Defever, Van den Bussche, & Reynvoet, 2011). Given the extent of the use of NDE-size in past literature as a metric of ANS acuity and the growing collection of recent publications that questions its validity, we believe it critical to examine the theoretical support as to whether NDE-size has the potential to serve as a metric of ANS acuity. Here, we model the theoretical relationship between ANS acuity and NDE-size. To foreshadow our conclusions, we do not find evidence that NDE-size can serve as a general measure of ANS acuity.

Individual Differences, ANS Acuity, and the NDE

Since Halberda et al.'s (2008) seminal work retrospectively predicting elementary school math ability from ANS acuity in fourteen-year-olds, a flurry of studies have linked ANS acuity to mathematical and academic success in children, adolescents, and adults (see Chen & Li, 2014, for meta-analysis and review). This connection between ANS acuity and other numerical skills can be explained by a mapping between non-symbolic ANS magnitudes and symbolic numbers. Humans, unlike other animals, have resources beyond the ANS to help them evaluate numbers. People can represent numbers verbally (e.g., "three thousand") and with other symbols (e.g., 3000) and do so with much greater precision than the ANS can achieve: The symbolic number 3000 is distinguished from 2999, but one cannot perceive the magnitude difference between 3000 and 2999 grains of rice. However, there is strong evidence that people invoke ANS-based analog magnitudes when considering symbolic numbers (Dehaene, Bossini, & Pascal, 1993; Moyer & Landaur, 1967). In Moyer and Landaur's (1967) seminal study, it was demonstrated that people show *distance effects* when making judgments about the quantities referenced by symbolic magnitudes. For example, people are faster at determining 6 is smaller than 9 than that 7 is smaller than 8. Such effects would result from neither use of symbolic look-up tables nor sequential-count based comparison processes, but (as discussed below) are a classic pattern in analog magnitude comparisons (Moyer & Landaur).

If symbolic number processing involves the ANS, one could therefore predict that performance on tasks involving symbolic numbers may be influenced by individual differences in ANS acuity. Moreover, as higher order mathematical skills build off one's understanding of symbolic numbers, one could further posit that these mathematical skills may be predicted by one's ANS acuity. Following this logic, some tasks used to assess individual differences in the ANS use symbolic numbers rather than non-symbolic numerical magnitudes (e.g., Holloway & Ansari, 2009; Seilger & Opfer, 2003; Sekuler & Mierkiews, 1977), even though symbolic numbers access ANS magnitudes only indirectly. Indeed, it appears that Moyer and Landaur's now classic approach of using the presence of distance effects to demonstrate that the ANS is invoked in symbolic magnitude comparisons inspired the later use of NDE-size as an ANS acuity metric (Sekuler & Mierkiews).

ANS Theory and NDE-Size

The exact nature of the ANS has yet to be completely determined, but it is well established that the perception of non-symbolic numerical magnitudes 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, 2009), numerical magnitudes are not perceived exactly, but rather percepts are normally or quasi-normally distributed around a mean value (which may itself be biased). The ability to distinguish between two quantities is dependent on the amount of overlap between their perceived magnitude distributions. Importantly, the overlap in the distributions of any two values - and thus the ease with which two values can be distinguished - is dependent upon their ratio, rather than upon the absolute distance between them. As a result, one can observe both *size effects* and *distance effects* in magnitude discriminations. 'Distance effects' refers to the observation that, within a given range, 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 [:::::]). 'Size effects' refers to the observation that it is easier to distinguish numerical quantities at the same distance in smaller magnitude ranges (6 dots [:::] vs. 8 dots [::::]) than in larger magnitude ranges (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. RTs similarly decrease with the comparison ratio (Whalen et al., 1999, see Kingdom & Prins, 2009, for a discussion of psychophysical functions).

To put it simply, Weber's law implies that a) the standard deviation (SD) of the distribution around an estimated magnitude is proportional to that magnitude's mean (M) and b) this proportion is constant. This constant proportion is, by definition, the Weber Fraction (w) of the perceiver's ANS. Thus, ANS acuity is defined by an individual's 'Weber Fraction' (Cordes et al., 2001; Deheane, et al., 2008; Halberda, Mazzocco, & Feigenson, 2008, Seigler & Opfer, 2003; Whalen et al., 1999). After accounting for other biases, this 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 non-symbolic magnitudes apart (ANS acuity). The smaller an individual's w, the better an individual will be at discriminating between non-symbolic numerical magnitudes, because there is less overlap in their numerical magnitude perceptions. Also, following Weber's Law, there is greater overlap in the magnitude distributions perceived from stimuli at smaller ratios (::::/:::, 1.33) than at larger ratios (::::::/:::, 2).

It follows that the ANS' contribution to an individual's NDE-size should be a function of the specific magnitudes being compared and the individual's ANS acuity (their Weber fraction, w). Thus, we can model the relative size of the ANS' contribution to NDE-size for any specific task and any given w by calculating the proportion of overlap of the distributions for any set of magnitude comparison pairs, and then calculating the savings in overlap for longer distances. For example, we can find the predicted ANS magnitude distribution overlap at short distances (4 vs. 5, 5 vs. 6) and at long distances (1 vs. 5, 5 vs. 9) for a given acuity level (w) and then find the differences between these overlaps. 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.

Model

As previously discussed, the ANS obeys Weber's Law. We follow the Linear Model of these phenomena here, which claims magnitudes perceived from stimuli are normally distributed, such that the means of perceived magnitude distributions increase linearly with the size of the stimuli and that the standard deviations of these distributions are proportional to their means (Cordes et al., 2001). Thus, the Coefficient of Variation (CV = SD/M) is constant for a given individual on a given task. This constant CV is the w of the individual's ANS: their ANS acuity. We model this here by representing perceived magnitudes as Gaussian distributions about unbiased means equal to the 'stimulus' value (M), where SD = w * M. As a result, the overlap in modeled ANS distributions is a function of the stimulus *ratio* and the Weber fraction (w) of the ANS: As SD is proportional to M, the overlap of the distributions derived from 8 and 10 is the same as that derived from 80 and 100. (Note that while the model presented in this manuscript treats magnitude representations as linearly spaced, with scalar variability, they might alternatively be modeled as logarithmically spaced with constant variability. However, a Logarithmic model would yield similar findings.)

Here, we calculated the proportion of overlap between modeled magnitude distributions. Additionally, following the method used by Halberda et al. (2008), we used the *erfc* (the complementary error function) to determine the rate at which a given pair of magnitudes will not be distinguished. Assuming there are no other sources of error (such as bias or inattention), this 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 equations used are given in the Appendix in the same Matlab code format we used for our calculations.

We seek to establish the theoretically maximal relationship between ANS and NDE-size, assuming no interfering factors or noise. Thus, we model theoretically ideal NDE-size at a given w for a given comparison pair as simply the differences between the overlaps or between the erfcs of different comparison ratios given that w. Real-world data would involve other sources of RT and error (attention to task, non-decision time, etc.) and make this relationship less clear. However, as these factors are separate from the ANS, they are excluded from this ideal model.

Results

The Relationship Between Overlap, ERFC, and *W*

In Figure 1, we illustrate the modeled ideal overlap of ANS magnitudes at various ratios (greater/lesser) as well as the calculated erfc. Recall, when controlling for total magnitude, smaller ratios map to smaller distances, and larger ratios map to larger distances. Smaller ratios have greater overlaps and greater error rates than larger ratios. The drop in both overlaps and erfcs is initially steep, but then "turns the corner" to asymptote to 0. Smaller ws (better acuity) and larger ws (worse acuity) both yield this same pattern, but the initial drop is steeper and the asymptote is reached faster for smaller ws. Using these values, the theoretical maximum contribution of ANS acuity to NDE-size can be found for any w on any particular NDE task.

The Relationship Between NDE-Size and *W*

As discussed above, an individual's NDE-size (i.e., savings in error rate and RT when comparing numerical magnitudes at large vs. small distances) should be functionally related to the difference in the amount of overlap between the comparison values' distributions, which is in turn related to the comparison ratios involved and the individual's w. We thus model NDE-size for a given w and pair of stimulus ratios as the difference in the model overlaps or erfcs:

NDE-size calculation for:
Overlap: Overlap(small ratio) – Overlap(large ratio)
erfc: erfc(small ratio) – erfc(large ratio)

We first model the ideal relationship between NDE-size and w for a comparison-pair set based on the distance effect paradigm originally developed by Moyer and Landauer (1967), and later expanded upon by researchers intending to measure NDE-size (see Sekuler & Mierkiews, 1977; Peters et al., 2008). In Moyer and Landaur's original paradigm, participants saw all possible pairs of non-equal integers between 1 and 9, and indicated which was larger. Although, it was Moyer and Landauer's intent simply to demonstrate that symbolic number comparisons yield distance effects, researchers often use modified forms of this procedure when measuring NDE-size (see Peters et al., 2008). A participant is shown a stimulus value and asked to indicate whether that value is greater or less than a central comparison value. The

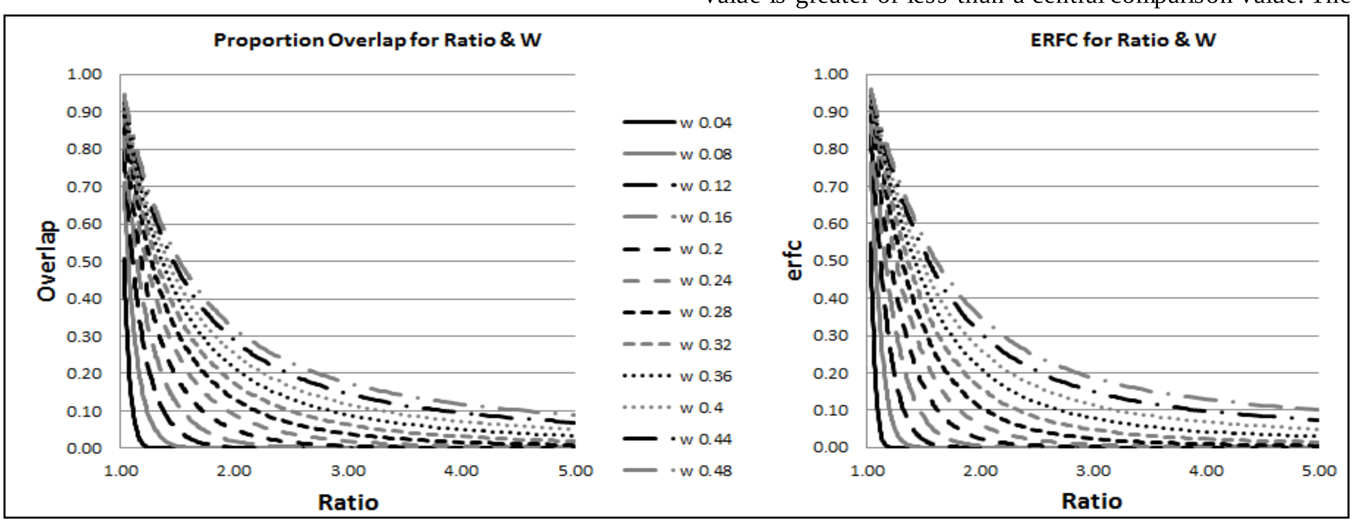

Figure 1: The modeled ANS magnitude distribution overlaps (left) and erfcs (right) for ws ranging from .04 (excellent acuity) to .48 (poorer acuity), at High/Low comparison value ratios ranging from 1 (equal values) to 5 (e.g., 50 vs. 10)

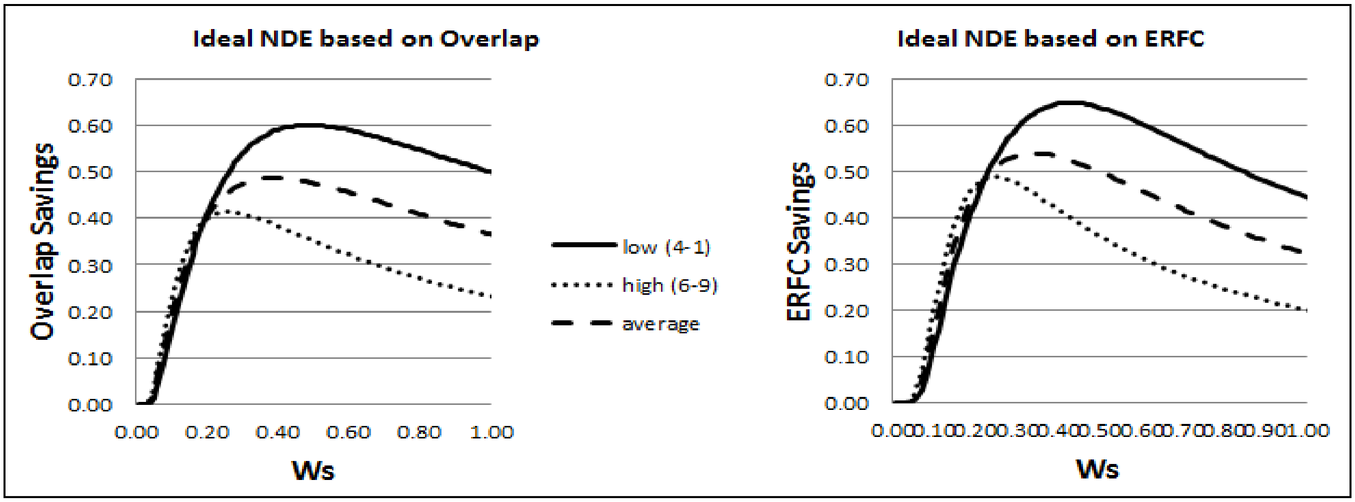

Figure 2. The modeled NDE-sizes for overlaps (left) and erfcs (right), based on values calculated for close distances (5 vs. 4 or 6) minus those calculated for far distances (5 vs. 1 or 9) distances for low ranges ([5 vs. 4] – [5 vs. 1]), high ranges ([6 vs. 5] – [9 vs. 5]), and their average.

stimuli follow a 2x2 design, varying both distance and direction from the central comparison value (e.g., 5). Half of the values 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 vs. far trials.

Recall, the overlap of ANS magnitude distributions (and thus the projected savings in RT and error rate) are dependent on the *ratio* of the values being compared, not their absolute distances. Thus, in this paradigm, although the absolute distances of the stimuli are symmetrical around the central comparison value, the ratios are asymmetrical. For the stimuli greater than the central value, the ratios are 6/5 (1.2) for the close value and 9/5 (1.8) for the far value. For the stimuli less than the central value, the ratios are 5/4(1.25) for the close values and 5/1 (5) for the far values. The spread of near and far ratios is thus much greater for the stimuli below than above the comparison value. However, the standard analyses used in the literature classify stimuli merely as near and far, and thus collapse across these ratio differences. We note that our model would yield the same NDE-sizes for any task using these same stimuli ratios and calculation methods regardless of the overall stimulus magnitudes invoked: The same pattern would be predicted for comparing 10, 40, 60, and 90 to a central value of 50.

The modeled NDE-sizes are presented in Figure 2. As can be seen, the relationship between *w* and NDE-size is not linear. Rather, it follows an inverted J shaped curve. NDE-size initially increases with *w*, slows to a peak, and then decreases with *w*. A strong positive linear relationship between *w* and NDE-size only exists for *w*s ranging between ~.05 and ~.20, quickly rising from near 0 savings to a 40% overlap savings, and 45% erfc savings. Both overlap and erfc then are near flat between *w*s of .20 and .60, with overlap savings peaking at 49% for *w*s of .39 and efrc savings peaking at 54% for *w*s of .34. Savings in overlap and erfc decline slowly with *w*s past these peaks. Clearly, the general presumption that larger NDE-sizes correlate with larger *w*s does not always hold. One could only expect to find a positive correlation with NDE-size and *w* if the target 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 non-existent correlation between NDE-size and *w*.

An alternative method of gauging NDE-size is to look at the slope of the regression of RTs or error rates on ratio or distance, treating distance in a continuous fashion, rather than dichotomizing it to 'close' and 'far' (see Sekuler & Mierkiewicz, 1977). Negative slopes indicate the presence of a distance effect as larger comparison ratios (higher/lower) and – within a given magnitude range – larger distances would typically yield faster RTs and fewer errors than smaller ratios and distances. "Larger" (i.e., more strongly negative) absolute slopes are treated as indicating larger *w*s, and thus poorer ANS acuity.

Here we model the theoretically ideal NDE-slopes one would predict based on both the ratios and the absolute distances used for the dot-array comparison task developed by Chesney, Bjälkebring, and Peters (2015). This task used ratios between 1 and 2.6, with the total number of dots in an array ranging between 10 and 30. As illustrated in Figure 3, the same pattern emerges here as with the dichotomized analysis: a J-shaped curve. A strong linear relationship between *w* and slope only holds between *w*s of .08 and .2. Moreover, this difference in overlap begins to decrease with *w* as *w* exceeds .32. Although the exact location of these inflection points will vary depending on the stimuli used in a particular study, those we show in Figure 3 align very closely to those we modeled for paradigms finding the "slope" of the NDE, where RTs (here, overlaps) are regressed on the distances between compared values for all possible unequal parings of the values 1-9 (see Sekuler & Mierkiewicz, 1977). We also demonstrate regressions using ratio as the IV yield a stronger relationship between NDE-slope and *w*, but in either case this relationship is non-linear.

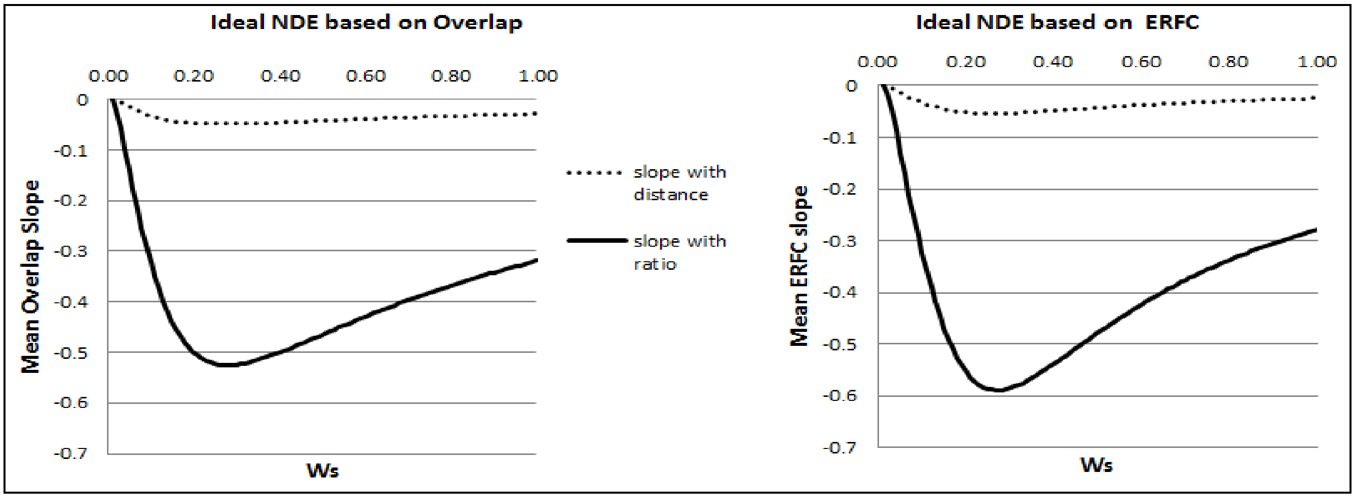

Figure 3: The modeled NDE-slopes for overlaps (left) and erfcs (right) for given *w*s based on both distances and ratios

Discussion

Despite the fact that the presence of numerical distance effects can indicate the involvement of the ANS in a task, it is becoming apparent that NDE tasks have limited utility for measuring individual differences in ANS acuity. With this model, we provide a novel theoretical exploration of the reasons why this is the case: Even under ideal conditions, one cannot expect NDE-size and ANS acuity to be linearly related. Rather, NDE-size initially increases, but then decreases with ANS acuity, with small NDE-sizes expected both for individuals with particularly small and particularly large *w*s. Our model demonstrates that the relationship between NDE-slope or NDE-size and *w* is dependent on both the ratio range of the stimuli and the *w* range in the population. This can be illustrated by considering the modeled NDE values in Figure 1. For example, if the target population's *w*s range between .04 and .20 and stimuli-pair ratios range between 1.25 and 2, one could indeed predict large NDE-sizes for larger *w*s (worse acuity), because the smaller *w*s would have already neared asymptote for this range of ratios, and would thus have small slopes. However, if the target population's *w*s ranged from .2-.4 and the stimuli ratios ranged from 1.05-1.2, the opposite pattern would emerge. These stimuli ratios are sufficiently small that all of these *w*s would be in the initial drop-off range, with smaller *w*s dropping faster: Smaller *w*s, (better acuity) would yield larger slopes.

For typical NDE-tasks, our model shows peak NDE-size is approached at *w*s of ~.2-.3. The location of this peak is a real concern, given the distribution of ANS *w*s in the normal population. Several studies with educated adult participants have found that their *w*s typically center around ~.22 (e.g., Cordes et. al, 2001; Whalen et al., 1999). Moreover, other studies of educated adults have found mean *w*s of .11 (Dehaene et al., 2008), and studies with infants have found *w*s of 1.0 (Xu & Spelke, 2000). Thus, one cannot expect that the range of *w*s for the population under test will coincide with the range of *w*s for which the relationship of *w* to NDE-size is quasi-linear on a standard task. Neither can researchers *a priori* assume a particular *w* distribution in a novel population, so as to be able to adjust their NDE tasks to yield a theoretically supported prediction of a linear relationship between NDE-size and *w*. 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.

Conclusions

Individual differences in ANS acuity have increasingly come to be considered an important predictor of human cognition and behavior. As such, it is important that the measures used to assess individual differences in ANS acuity are both reliable and valid. We recommend that future researchers assess ANS acuity via tasks whose reliability has been established. We specifically recommend against using NDE tasks to assess ANS acuity, as the validity of such measures is in doubt.

References

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. doi: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.* doi: 10.3758/s13414-015-0974-6

Cokely, E. T., Galesic, M., Schulz, E., Ghazal, S., & Garcia-Retamero, R. (2012). Measuring risk literacy: The Berlin numeracy test. *Judgment and Decision Making, 7*, 25- 47.

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. doi:10.3758/BF03196206

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. doi:10.1126/science.1156540
Dehaene, S., Bossini, S. & Pascal, G. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General. 122*, 371-396.
Gilmore, C., Attridge, N., & Inglis, M. (2011). Measuring the approximate number system. *The Quarterly Journal of Experimental Psychology, 64,* 2009-2109. doi: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.
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. doi:10.1037/a0012682
Halberda, J., Mazzocco, M. & Feigenson, L. (2008). Individual differences in nonverbal number acuity predict maths achievement. *Nature, 455,* 665-668. doi: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. doi:10.1016/j.jecp.2008.04.001.
Ingis, M. & Gilmore, C. (2014). Indexing the approximate number system. *Acta Psychologica, 145*, 147-155. doi:10.1016/j.actpsy.2013.11.009
Lipkus, I. M., Samsa, G., & Rimer, B. K. (2001). General performance on a numeracy scale among highly educated samples. *Medical Decision Making, 21*, 37–44. doi:10.1177/0272989X0102100105
Kaufman, E. L., Lord, M. W., Reese, T. W., & Volkmann, J. (1949). The discrimination of visual number. *American Journal of Psychology, 62*, 498-535.
Kingdom, F.A.A. & Prins, N. (2010). *Psychophysics: A Practical Introduction*. Academic Press: an imprint of Elsevier, London.
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. doi: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. doi: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. doi: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. doi:10.1038/2151519a0
Peters, E., Slovic, P., Västfjäll, D., & Mertz, C. K. (2008). Intuitive numbers guide decisions. Judgment and Decision Making, 3, 619-635. Available at SSRN: http://ssrn.com/abstract=1321907
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.
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.
Schley, D.R., & Peters, E. (2014). Assessing "economic value": Symbolic number mappings predict risky and riskless valuations. *Psychological Science, 25*, 753-761. doi:10.1177/0956797613515485
Sekular, R., & Mierkiewiez, D. (1977). Children's Judgments of Numerical Inequality. *Child Development, 48*, 630-633. doi: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. doi:10.1111/1467-9280.02438
Taves, E. H. (1941). Two mechanisms for the perception of visual numerousness. *Archives of Psychology, 37*(265).
Weller, J., Dieckmann, N. F., Tusler, M., Mertz, C. K., Burns, W., & Peters, E. (2013). Development and testing of an abbreviated numeracy scale: A Rasch Analysis approach. *Journal of Behavioral Decision Making, 26*, 198-212. doi: 10.1002/bdm.1751
Whalen, J., Gallistel, C. R., & Gelman, R. (1999). Non-verbal counting in humans: The psychophysics of number representation. *Psychological Science, 10,* 130-137. doi:10.1111/1467-9280.00120
Xu, F., & Spelke, E.S. (2000). Large number discrimination in 6-month-old infants. *Cognition, 74,* B1-B11.

Acknowledgments

This research was supported by NSF grant SES–1155924. We thank Ellen Peters for her help and support.

Appendix

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

```
invCdf = 1- normcdf((High - Low)/(w*(High+Low)),0,1)
overlap = (2*invCdf)/(2-(2*invCdf))
```

The Matlab code used to find the erfc – the value of the complimentary error function – which is the rate at which the ANS values will not be distinguished and thus double the ideal error rate:

```
erfc(abs(High-Low)/(sqrt((High^2)+(Low^2))*sqrt(2)*w))
```

"High" referred to the higher value in a comparison pair (e.g., 6), while "Low" referred to the lower value in the pair (e.g., 5). "w" referred to ANS acuity (Weber fraction, *w*).