Commentary on:
Leibovich, Katzin, Harel & Henik
"From 'sense of number' to 'sense of magnitude' – The role of continuous magnitudes in numerical cognition"

Abstract word count: 60
Main text word count: 966
References word count: 511
Entire text word count: 1640

Title:
Is the ANS linked to mathematics performance?

Authors:
Matthew Inglis (Loughborough University)
Sophie Batchelor (Loughborough University)
Camilla Gilmore (Loughborough University)
Derrick G. Watson (University of Warwick)

Address correspondence to:
Matthew Inglis, Mathematics Education Centre, Loughborough University, Loughborough, Leicestershire. LE11 3TU. United Kingdom.

Emails:
m.j.inglis@lboro.ac.uk
s.m.batchelor@lboro.ac.uk
c.gilmore@lboro.ac.uk
d.g.watson@warwick.ac.uk

Homepages:
http://homepages.lboro.ac.uk/~mamji/
http://www.lboro.ac.uk/departments/mec/staff/sophie-batchelor.html
http://www.lboro.ac.uk/departments/mec/staff/camilla-gilmore.html
https://www2.warwick.ac.uk/fac/sci/psych/people/dwatson/dwatson/

Abstract (60 words)

Leibovich, Katzin, Harel, & Henik argue persuasively that researchers should not assume ANS tasks harness an innate sense of number. However, some studies have reported a causal link between ANS tasks and mathematics performance, implicating the ANS in the development of numerical skills. Here we report a *p*-curve analysis which indicates that these experimental studies do not contain evidential value.

Main text (1000 words)

As Leibovich, Katzin, Harel, & Henik (LKH&H) point out, the dominant view is that mechanisms involved in performing 'number sense' or Approximate Number System (ANS) tasks underlie the basis of symbolic mathematical skill. This view is based on findings from two main experimental paradigms; in comparison tasks participants select which of two arrays contain more dots, in addition tasks they assess whether two sequentially displayed arrays contain more dots than a third array. Researchers have assumed that such ANS tasks harness an innate sense of number, but LKH&H argue that this assumption is not warranted.

In our view, there are three main sources of evidence for the view critiqued by LKH&H:

1. Face validity. Tasks in which children or adults compare, for example, the number of yellow and blue dots do, on the face of it, seem to be about number.
2. Correlational evidence. Recent meta-analyses have reported that performance on standardized mathematics tests and ANS tasks correlate at $r = 0.2$ to 0.3 (Chen & Li, 2014; Fazio, Bailey, Thompson & Siegler, 2014; Schneider, Beeres, Merz, Schmidt, Stricker, & De Smedt, 2016).
3. Causal evidence. Some recent experimental studies have claimed that improving performance on ANS tasks causes higher mathematics achievement and faster mathematics performance.

LKH&H argue compellingly that evidence from comparison tasks is insufficient to conclude that ANS tasks involve numerical processing, at least not as currently conceived by proponents of the ‘number sense’ theory (cf. Gebuis, Cohen Kadosh & Gevers, 2016). Further, accounting for the ANS/mathematics achievement correlation does not require the assumption of an innate sense of number, because of the inhibitory control demands of incongruent trials on ANS tasks (e.g., Fuhs & McNeil, 2013; Gilmore et al., 2013). However, this inhibition confound does not account for the third source of evidence, which LKH&H do not address.

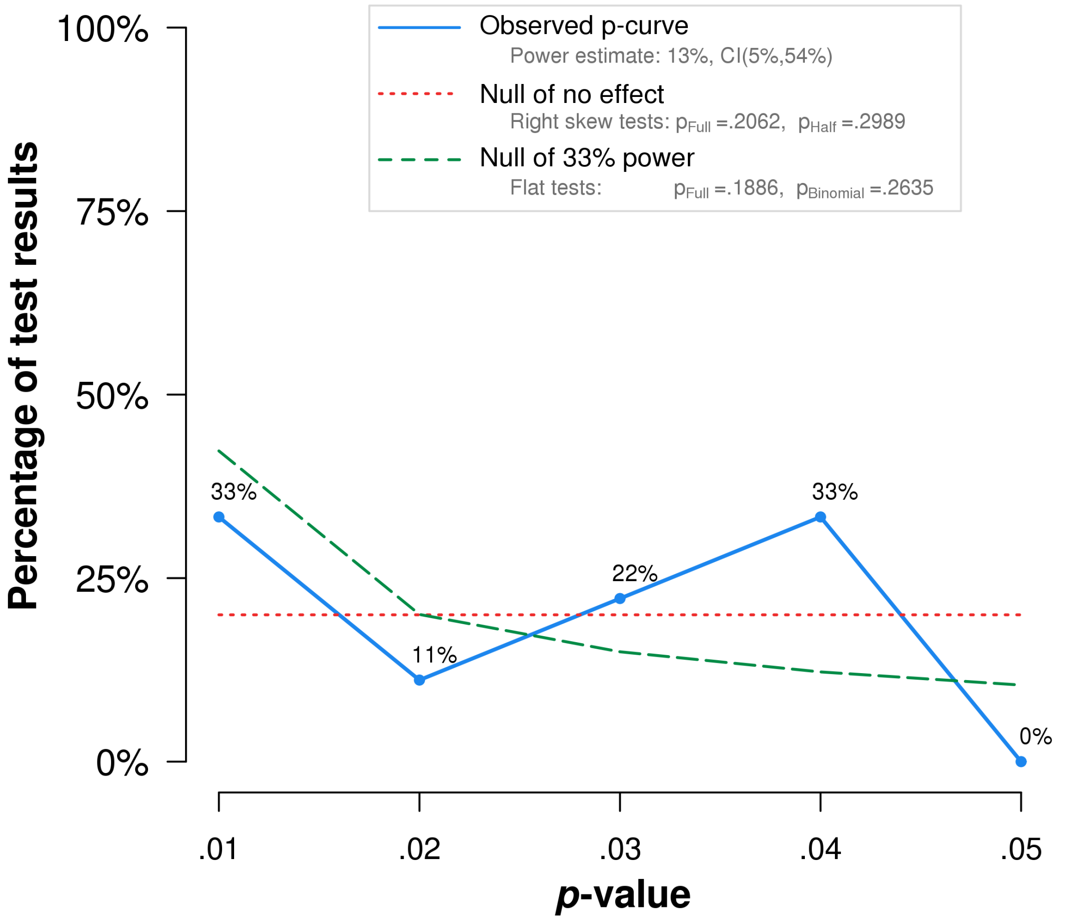
The causal evidence comes from two sources. One line of research has found training on ANS tasks leads to improved performance and faster responses on mathematics tests (Hyde, Khanum & Spelke, 2014; Khanum, Hanif, Spelke, Berteletti, & Hyde, 2016; Park, Bermudez, Roberts & Brannon, 2016; Park & Brannon, 2013, 2014). Another has found that manipulating the order in which ANS trials are presented (easy-to-difficult or difficult-to-easy) improves mathematics performance (Wang, Odic, Halberda & Feigenson, 2016). These findings present a problem for the inhibition account. Earlier research found that inhibition training does not transfer to non-trained tasks (Thorell, Lindqvist, Bergman, Bohlin, & Klingberg, 2008) so potential inhibitory control demands of ANS tasks cannot explain these findings.

While there has been a debate about the quality of this evidence (e.g., Lindskog & Winman, 2016; Merkley, Matejko, & Ansari, 2017; Park & Brannon, 2016; Wang, Odic, Halberda, & Feigenson, 2017), here we ask whether, taken together at face value, current experimental studies provide sufficient evidence to conclude that there is a causal link between the ANS and mathematics performance. To this end, we performed a *p*-curve analysis on the set of all studies we are aware of that report a causal link between the ANS and mathematics performance (Hyde et al., 2014; Khanum et al., 2016; Park et al., 2016; Park & Brannon, 2013, 2014; Wang et al., 2016).

P-curve analyses (Simonsohn, Nelson, & Simmons, 2014; Simonsohn, Simmons, & Nelson, 2015) rely on the fact that *p*-values follow a uniform distribution under the null hypothesis. In contrast, when the null is false *p*-values are right skewed (i.e. there are more low values than high values). This is true for the full 0 to 1 interval, but also for the interval from 0 to 0.05. Simonsohn et al. proposed that the shape of the distribution of significant *p*-values in a set of studies, can be used to assess if they collectively contain evidential value. If the significant *p*-values follow a roughly uniform distribution, publication bias might explain the results.

We followed Simonsohn et al.'s (2015) procedure and, for each reported study, extracted the test statistic associated with the hypothesis of interest (whether the experimental manipulation influenced mathematics performance). If there was doubt about which statistic to select (e.g., the study contained two control groups), we conservatively selected the comparison with the smaller *p*-value (retaining the other for a robustness check). Details are given in our *p*-curve disclosure table at https://dx.doi.org/10.6084/m9.figshare.4262999.v1.

We analyzed the test statistics using the *p*-curve app v4.05 (http://www.p-curve.com/app4/). The *p*-value distribution is shown in Figure 1. Of 9 *p*-values, 5 were below .025, a frequency not significantly different to the 4.5 expected under the null hypothesis, one-tailed binomial test, p=.5. Stouffer's method (Simonsohn et al., 2015) also indicated that these studies do not contain evidential value (*p*s = .206, .299). The *p*-curve method also provides an estimate of the power of the studies. Here this was 13%, 90% CI [5%, 54%], indicating insufficient evidence to reject the null of 33% power (which Simonsohn et al. would take to indicate evidential value was absent and that replications would not be expected to succeed).



Note: The observed p-curve includes 9 statistically significant (p<.05) results, of which 5 are p<.025. There were no non-significant results entered.

Figure 1. The distribution of *p*-values for studies finding a causal connection between the ANS and mathematics performance.

Our findings indicate that the published literature to date does not contain evidence of a causal link between performance on ANS tasks and standardized mathematics tests. To be clear, we have not demonstrated there is no causal connection between the ANS and mathematics performance, only that the existing literature does not provide evidence for one. However, we can definitively conclude that existing studies are substantially underpowered, rendering their interpretation difficult. In future researchers should address this limitation through preregistration and larger samples.

To conclude, we endorse LKH&H's suggestion that the assumption ANS tasks involve number sense is not justified. Although LKH&H did not address it, we believe that existing evidence of a causal link between the ANS and mathematics performance is insufficient to challenge their argument.

Reference List

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.

Gebuis, T., Cohen Kadosh, R., & Gevers, W. (2016). Sensory-integration system rather than approximate number system underlies numerosity processing: A critical review. *Acta Psychologica, 171,* 17-35.

Gilmore, C., Attridge, N., Clayton, S., Cragg, L., Johnson, S., Marlow, N., …, Inglis, M. (2013). Individual differences in inhibitory control, not non-verbal number acuity, correlate with mathematics achievement. *PLOS ONE, 8*, e67374.

Fazio, L. K., Bailey, D. H., Thompson, C. A., & Siegler, R. S. (2014). Relations of different types of numerical magnitude representations to each other and to mathematics achievement. *Journal of Experimental Child Psychology, 123*, 53–72.

Hyde, D. C., Khanum, S., & Spelke, E. S. (2014). Brief non-symbolic, approximate number practice enhances subsequent exact symbolic arithmetic in children. *Cognition, 131*, 92-107.

Khanum, S., Hanif, R., Spelke, E., Berteletti, I. & Hyde, D. C. (2016). Effects of non-symbolic approximate number practice on symbolic numerical abilities in Pakistani children. *PLOS ONE, 11*, e0164436

Lindskog, M. & Winman, A. (2016). No evidence of learning in non-symbolic numerical tasks – A comment on Park & Brannon (2014). *Cognition, 150,* 243-247.

Merkley, R., Matejko, A. A., & Ansari, D. (2017). Strong causal claims require strong evidence: A commentary on Wang and colleagues. *Journal of Experimental Child Psychology, 153,* 163-167.

Park, J., Bermudez, V., Roberts, R. C., & Brannon, E. M. (2016). Non-symbolic approximate arithmetic training improves math performance in preschoolers. *Journal of Experimental Child Psychology, 152*, 278-293.

Park, J. & Brannon, E. M. (2013). Training the Approximate Number System Improves Math Proficiency. *Psychological Science, 24*, 2013-2019.

Park, J. & Brannon, E. M. (2013). Improving arithmetic performance with number sense training: An investigation of underlying mechanism. *Cognition, 133*, 188-200.

Park, J. & Brannon, E. M. (2016). How to interpret cognitive training studies: A reply to Lindskog & Winman. *Cognition, 150,* 247-251.

Schneider, M., Beeres, K., Coban, L., Merz, S., Schmidt, S., Stricker, J., & De Smedt, B. (in press). Associations of non-symbolic and symbolic numerical magnitude processing with mathematical competence: A meta-analysis. *Developmental Science.* doi: 10.1111/desc.12372

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). *P*-curve: A key to the file-drawer. *Journal of Experimental Psychology: General, 143,* 534-547.

Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2015). Better P-Curves: Making P-Curve Analysis More Robust To Errors, Fraud, and Ambitious P-Hacking, A Reply To Ulrich and Miller (2015). *Journal of Experimental Psychology: General, 144*, 1146-1152.

Thorell, L. B., Lindqvist, S., Bergman Nutley, S., Bohlin, G., & Klingberg, T. (2009). Training and transfer effects of executive functions in preschool children. *Developmental Science, 12*, 106–113.

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.

Wang, J., Odic, D., Halberda, J., & Feigenson, L. (2017). Better together: Multiple lines of evidence for a link between approximate and exact number representations: A reply to Merkley, Matejko, and Ansari. *Journal of Experimental Child Psychology, 153,* 168-172.