Acta Psychologica 145 (2014) 147–155

Contents lists available at ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/ locate/actpsy

Indexing the approximate number system☆

Matthew Inglis *,[1], Camilla Gilmore

Mathematics Education Centre, Loughborough University, United Kingdom

ARTICLE INFO

Article history:
Received 15 March 2013
Received in revised form 26 November 2013
Accepted 27 November 2013
Available online 17 December 2013

PsycINFO classification:
2340
2343

Keywords:
Numerical cognition
Mathematics
Approximate number system

ABSTRACT

Much recent research attention has focused on understanding individual differences in the approximate number system, a cognitive system believed to underlie human mathematical competence. To date researchers have used four main indices of ANS acuity, and have typically assumed that they measure similar properties. Here we report a study which questions this assumption. We demonstrate that the numerical ratio effect has poor test–retest reliability and that it does not relate to either Weber fractions or accuracy on nonsymbolic comparison tasks. Furthermore, we show that Weber fractions follow a strongly skewed distribution and that they have lower test–retest reliability than a simple accuracy measure. We conclude by arguing that in the future researchers interested in indexing individual differences in ANS acuity should use accuracy figures, not Weber fractions or numerical ratio effects.

1. Introduction

How do students develop their mathematical competence? In recent years there has been substantial interest in addressing this question by investigating individual differences in children and adults' abilities when performing basic arithmetic operations on nonsymbolic stimuli. Infants, children, adults and non-human animals are all capable of forming rapid nonsymbolic representations of the numerosity of arrays of dots and sequences of tones (e.g., Cordes, Gelman, Gallistel, & Whalen, 2001; Dehaene, 1997; Feigenson, Dehaene, & Spelke, 2004). The mechanism that underlies these representations has become known as the approximate number system (or ANS) and allows individuals to compare, add, and subtract sets of items, e.g., objects, dots, or tones (Barth, La Mont, Lipton, Dehaene, Kanwisher & Spelke, 2006; Meck & Church, 1983; Pica, Lemer, Izard, & Dehaene, 2004).

Some researchers have hypothesised that the ANS is the cognitive basis of all formal symbolic mathematics abilities; several sources of evidence support this view. First, the ANS is automatically activated in response to Arabic numerals in addition to nonsymbolic arrays (Moyer & Landauer, 1967). Second, prior to formal mathematical instruction children seem to be capable of using ANS mechanisms to perform approximate calculations with Arabic numerals despite being incapable of performing exact calculations (Gilmore, McCarthy, & Spelke, 2007). Third, measures of the precision of children's ANS representations – their so-called ANS acuity – have been found in some studies to predict their achievement on standardised school mathematics tests (e.g., De Smedt, Verschaffel & Ghesquière, 2009; Halberda, Mazzocco, & Feigenson, 2008; Inglis, Attridge, Batchelor, & Gilmore, 2011; Libertus, Feigenson, & Halberda, 2011; Mazzocco, Feigenson, & Halberda, 2011a; Mundy & Gilmore, 2009). Fourth, it has been found in some studies that students with dyscalculia have lower ANS precision than typically achieving children, suggesting that an ANS deficit may be the cause of mathematical learning difficulties (Mazzocco, Feigenson & Halberda 2011b; Piazza, Facoetti, Trussardi, Berteletti, Conte, Lucangeli, Dehaene & Zorzi, 2010).

All these studies rely upon measuring an individual's ANS acuity: the accuracy with which they represent nonsymbolic numerosities. Typically this is achieved using the nonsymbolic comparison task. Participants are presented with two dot arrays n_1 and n_2, side by side or sequentially, and asked to judge which is the larger. After the presentation of many such pairs, one of four indices is typically calculated: accuracy, Weber fraction, numerical ratio effect (NRE) for accuracy or NRE for reaction time. These four indices are implicitly assumed to be measuring the same property: the acuity of an individual's ANS (e.g. Libertus, Odic & Halberda, 2012; Price, Palmer, Battista, & Ansari, 2012). But, to date, little evidence has been presented for this suggestion. Our goal in this paper is to investigate the psychometric properties of, and interrelations between, these different indices. Before motivating our specific questions, we briefly discuss each of the four indices.

Several researchers have, when investigating ANS acuity, simply reported participants' accuracies: the proportion of trials they answered correctly (e.g., Fuhs & McNeil, 2013; Gilmore, Attridge, & Inglis, 2011; Kolkman, Kroesbergen, & Leseman, 2013; Lourenco, Bonny, Fernandez, & Rao, 2012; Nys, Ventura, Fernandes, Querido, Leybaert, & Content,

☆ This research was supported by a British Academy Postdoctoral Fellowship (C.G.), and a Royal Society Worshipful Company of Actuaries Educational Research Fellowship (M.I.).

* Corresponding author.
E-mail address: m.j.inglis@lboro.ac.uk (M. Inglis).
[1] Mathematics Education Centre, Loughborough University, Loughborough, Leicestershire, LE11 3TU, United Kingdom.

0001-6918/$ – see front matter
http://dx.doi.org/10.1016/j.actpsy.2013.11.009

2013; Wei, Yuan, Chen & Zhou, 2011) or, less commonly, the number of trials they answered correctly in a given time (e.g., Nosworthy, Bugden, Archibald, Evans & Ansari, 2013).

The Weber fraction is an alternative approach to indexing an individual's ANS acuity (e.g. Bonny & Lourenco, 2013; Castronovo & Göbel, 2012; Halberda & Feigenson, 2008; Halberda, Ly, Willmer, Naiman, & Germine, 2012; Halberda et al., 2008; Inglis et al., 2011; Libertus et al., 2011, 2012; Lyons & Beilock, 2011; Mazzocco et al., 2011a; Piazza et al., 2010; Price et al., 2012; Sasanguie, Gobel, Moll, Smets & Reynvoet, 2013). It makes the theoretical assumption that the ANS operates according to the Weber–Fechner law (e.g. Barth et al., 2006). Under this interpretation, when an individual observes an array of n dots, they form an internal representation which follows a normal distribution with mean n and standard deviation wn. Here w is the Weber fraction, which represents the precision of the individuals' representation. Those with ws closer to zero are more likely to form representations closer to the true value of the numerosity n. These assumptions imply that an individuals' expected accuracy on a given trial is a function of n_1, n_2 and w: $\mathrm{acc}(n_1, n_2; w) = \frac{1}{2} + \frac{1}{2}\mathrm{erf}\left(\frac{|n_1 - n_2|}{\sqrt{2}w\sqrt{n_1{}^2 + n_2^2}}\right)$. In practice, an individual's Weber fraction can be estimated by calculating the value of w which bests fits their behavioural data.

Fig. 1 shows the relationship between the ratio of the two to-be-compared numerosities, and an individuals' expected accuracy for various values of w. As can be seen, as expected accuracy tends to 0.5, w asymptotically tends to infinity. It is therefore impossible for an individual to have an accuracy of under 0.5 under this model (to do so would require a w greater than infinity). The practical consequence of these considerations is that Weber fractions can only be calculated for participants whose responses follow the Weber–Fechner law, and consequently who score above 0.5 (cf. Libertus et al., 2011).

Finally, some researchers have adopted the numerical ratio effect (NRE), or the closely related numerical distance effect (NDE), to index ANS acuity (e.g., Bugden, Price, McLean & Ansari, 2012; Gilmore et al., 2011; Holloway & Ansari, 2009; Lonnemann, Linkersdorfer, Hasselhorn & Lindberg, 2011; Merkley & Ansari, 2010; Price et al., 2012; Sasanguie, Van den Bussche, & Reynvoet, 2012; Vanbinst, Ghesquière & De Smedt, 2012). This effect observes that individuals are typically less accurate on, and slower to respond to, comparison trials where the n_1/n_2 ratio is close to 1. An individual's NRE can be obtained by calculating the slope of their ratio–accuracy (or ratio–RT) graph. Assuming that $n_1/n_2 < 1$, then an individual with a strongly negative NRE(accuracy) shows a substantial drop off in accuracy between easier trials (with ratios away from 1) and harder trials (those with ratios close to 1). Similarly, an individual with a strongly positive NRE(RT) shows a substantial slowing between easier and harder trials. As Fig. 1 illustrates, the slope of an individuals' ratio–accuracy curve is predicted by their Weber fraction (the slope of the $w = 0.1$ curve is substantially steeper than the slope of the $w = 0.4$ curve, for example). Therefore, an individuals' NRE(accuracy) should, according to the standard model of the ANS, be strongly related to their Weber fraction (albeit non-linearly). It is less clear whether theory would predict a relationship between Weber fractions and NRE(RT)s, although many researchers have used NRE(RT) to index ANS acuity (e.g. Price et al., 2012).

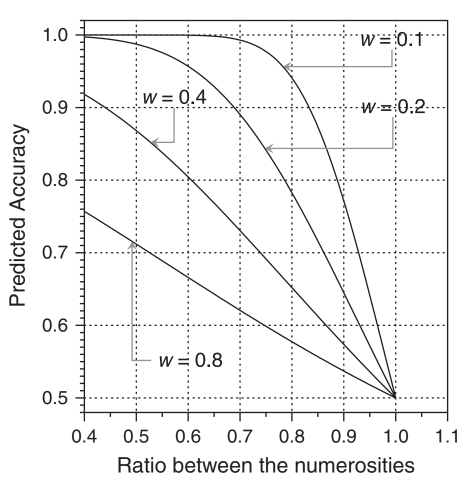

Fig. 1. Predicted accuracy as a function of the n_1/n_2 ratio, for various values of w.

The four different methods of indexing ANS acuity have, to a large extent, been assumed to unproblematically measure the same phenomenon (e.g., Libertus et al., 2012; Price et al., 2012). However, there are at least four reasons to doubt this belief.

First, calculations of the reliability of the different indices have been surprisingly low. Price et al. (2012) calculated immediate test–retest reliability figures for the NRE(RT) and Weber fraction on three variants of the nonsymbolic comparison task, finding reliability coefficients varying between $r = .4$ and .8; Maloney, Risko, Preston, Ansari and Fugelsang (2010) found that the immediate test–retest reliability of an NDE(accuracy) measure was in the same range, $r \approx .6$. Remarkably, Libertus et al. (2012) found that the three month test–retest reliability of their measure of individuals' Weber fractions was not significantly different to zero.

Second, researchers have observed surprisingly low correlations between estimates of these indices obtained from different tasks. For example, Gilmore et al. (2011) found that estimates of Weber fraction obtained from a nonsymbolic comparison task did not correlate with similar indices derived from a nonsymbolic addition task, which is believed to be a closely related method of assessing ANS acuity (e.g. Barth et al., 2006).

Third, researchers have reported different relationships between their measures of individuals' ANS acuity and mathematical achievement while some of these researchers have indexed ANS acuity using Weber fractions (e.g. Castronovo & Göbel, 2012; Halberda & Feigenson, 2008, Halberda, Ly, Willmer, Naiman and Germine, 2012; Halberda et al., 2008; Inglis et al., 2011; Libertus et al., 2012; Lyons & Beilock, 2011; Piazza et al., 2010; Price et al., 2012; Sasanguie et al., 2012), others have used NREs (e.g., Bugden et al., 2012; Holloway & Ansari, 2009; Lourenco et al., 2012; Merkley & Ansari, 2010; Price et al., 2012), and others accuracy (e.g. Fuhs & McNeil, 2013; Nys et al., 2013; Wei, Yuan, Chen and Zhou, 2011). One account for why some of these researchers have found a relationship between ANS acuity and mathematics achievement, and others have not, is simply that their choice of index does not measure the same underlying phenomenon. For example, Mundy and Gilmore (2009) found a significant relationship between nonsymbolic comparison performance and mathematical achievement, but only when they indexed performance by accuracy rather than NDE.

Finally, to our knowledge, the only attempt to understand the relationship between different indices of ANS acuity suggests that the indices may measure different phenomena. Price et al. (2012) found extremely weak relationships between NDE(RT)s and Weber fractions. They found a significant (but weak, $R^2 = .11$) association between these two indices on a nonsymbolic comparison task where the stimuli were presented sequentially, and no significant associations on tasks where the stimuli were displayed concurrently.

To summarise, although much progress has been made towards understanding the ANS and its relationship with mathematical achievement, there is little agreement in the literature about how best to index an individual's ANS acuity. Further, there are reasons to suppose that at least some contradictory findings reported in the literature could be resolved by a careful study of the psychometric properties of different indices of the ANS. In this paper we take a step in this direction by asking four main questions. First, what distributions do the four commonly-used indices of ANS acuity (accuracy, Weber fraction, NRE(accuracy) and NRE(RT)) follow? Second, what are the relationships between these different indices? Third, what are the immediate and delayed test–retest reliabilities of the different indices? Finally, to what extent are accuracies and Weber fractions dependent on the problem sets