Learning to focus on number

Manuela Piazza a,b,c,*, Vito De Feo d, Stefano Panzeri d, Stanislas Dehaene a,b,e

a Cognitive Neuroimaging Unit, CEA DSV/I2BM, INSERM, Université Paris-Sud, Université Paris-Saclay, NeuroSpin Center, 91191 Gif/Yvette, France

b NeuroSpin Center, Institute of BioImaging, Commissariat à l'Energie Atomique, F-91191 Gif/Yvette, France

c Center for Mind/Brain Sciences, University of Trento, 38068 Rovereto, Italy

d Center for Neuroscience and Cognitive Systems @UniTn, Neural Computation Laboratory, Istituto Italiano di Tecnologia, 38068 Rovereto, Italy

e Collège de France, F-75005 Paris, France

A B S T R A C T

With age and education, children become increasingly accurate in processing numerosity. This developmental trend is often interpreted as a progressive refinement of the mental representation of number. Here we provide empirical and theoretical support for an alternative possibility, the filtering hypothesis, which proposes that development primarily affects the ability to focus on the relevant dimension of number and to avoid interference from irrelevant but often co-varying quantitative dimensions. Data from the same numerical comparison task in adults and children of various levels of numeracy, including Mundurucú Indians and western dyscalculics, show that, as predicted by the filtering hypothesis, age and education primarily increase the ability to focus on number and filter out potentially interfering information on the non-numerical dimensions. These findings can be cap- tured by a minimal computational model where learning consists in the training of a multivariate classifier whose discrimination boundaries get progressively aligned to the task-relevant dimension of number. This view of development has important consequences for education.

1. Introduction

During development, children become increasingly precise in making numerical judgments (Halberda & Feigenson, 2008). The evi- dence for this change comes primarily from numerosity comparison or discrimination tasks, where participants are asked to point, without counting, to the numerically larger (or smaller) of two sets, or to decide whether two sets contain the same number of items. Performance on such tasks depends on the logarithm of the ratio (log ratio) of the two numerosities, according to Weber's law (Dehaene, 2007). Studies in naïve non-human animals (Agrillo, Dadda, Serena, & Bisazza, 2008; Jordan, Brannon, Logothetis, & Ghazanfar, 2005; Rugani, Regolin, & Vallortigara, 2011; Viswanathan & Nieder, 2015) and human newborns (Izard, Sann, Spelke, & Streri, 2009) indicate that number, like many numerical discrimination performance later on (Guillaume, Nys, Mussolin, & Content, 2013; Piazza, Pica, Izard, Spelke, & Dehaene, 2013; Nys et al., 2013).

The most straightforward explanation for this behavioral improve- ment, hereafter called the sharpening hypothesis, assumes that ma- turation and formal education progressively sharpen the internal re- presentation of numerosity, see Fig. 1B. The intraparietal cortex of both humans and macaques has been identified as a key node for the neural representation of numerosity (Piazza & Eger, 2015), and this hypothesis holds that the tuning curves of neurons in this region get progressively sharper. This idea recently received partial support by two fMRI studies investigating numerosity coding precision in the intraparietal sulcus (hereafter IPS) of adults and young preschoolers tested with an iden- tical adaptation paradigm: the pattern of fMRI responses to numerically other quantitative dimensions of the environment, is immediately deviant stimuli, a proXy for "numerosity tuning functions", were available, even in the absence of training. However, the precision of numerical discrimination is initially low (newborns discriminate sets only when they differ by 300%), and it improves progressively during development (adults eventually differentiate small 15–20% numerical changes) (Halberda & Feigenson, 2008). Recent investigations indicate that while brain maturation is responsible for this evolution during the first years of life, formal education plays a key role in increasing sharper in adults (Piazza, Izard, Pinel, Le Bihan, & Dehaene, 2004) compared to preschoolers (Kersey & Cantlon, 2017), mirroring their higher accuracy in numerical discrimination precision. However, be- cause the BOLD signal has limited temporal resolution, it remains possible that the brain activation in this paradigm reflected the effect of a post-perceptual attentional amplification rather than the initial en- coding of numerosity.

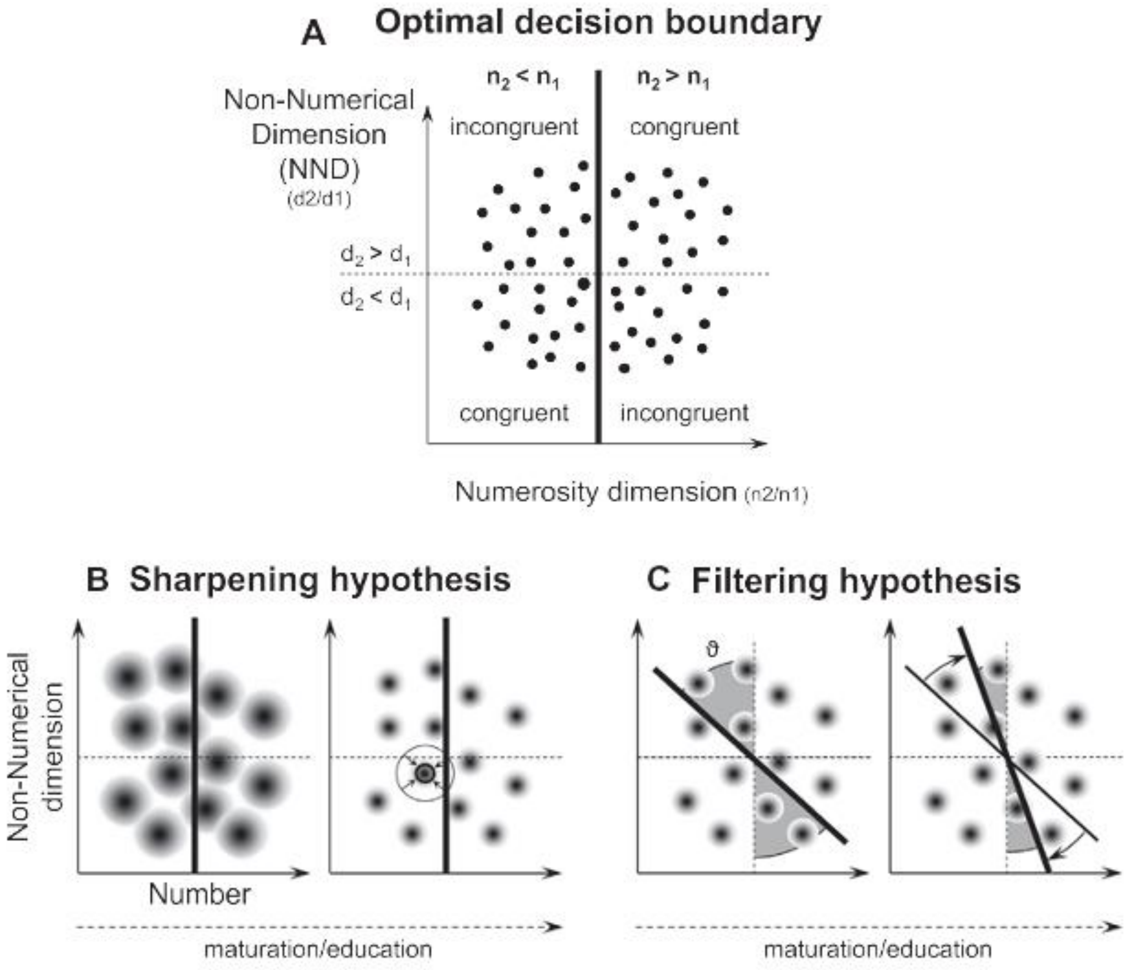


Fig. 1. Two potential sources of errors and two theoretical accounts of the increasing precision of numerical abilities with age and education. The task studied in the present paper requires com- paring pair of sets of different numbers (n1 and n2) each characterized by different non-numer- ical dimensions (such as total

occupied area or individual items size, d1 and d2), and to choose the set with the largest number ignoring the non- numerical dimensions. In panel A, each dot re- presents one individual trial that is a specific combination of two numbers (their log ratio varying along the horizontal axis) and their re- lative non-numerical features (their log ratio varying along the vertical axis). The vertical line indicates the optimal decision boundary for such a number comparison task. Dots close to the de- cision boundary represent trials where the two numerosities vary little, and their distance in- creases as we move away from it. Panel B and C represent the two potential sources of errors and of developmental changes, referred to as the "sharpening" and the "filtering" hypotheses respectively. In these panels the width of the dots represents the noise of internal representations of number and of another non-numerical dimension (NND). According to the sharpening model (B), representations are initially highly noisy and they become more precise (sharper) with age and education. Sharpening predicts an overall re- duction in error rates, particularly for stimuli close to the decision boundary, but not necessarily a reduction in the congruity effect: error rate should equally decrease in the congruent and incongruent pairs. According to the filtering model (C), numerical development involves an increasing capacity to focus on the relevant dimension and to filter out irrelevant non-numerical dimensions, with no concurrent change in the precision of the underlying representations. Such a development is illustrated here as a progressive rotation of the decision boundary towards the optimal vertical line, thus a reduction of the angle (ϑ) between the actual decision slope and the optimal one. Filtering predicts a reduction of the congruity effect in that error rates should solely decrease in the incongruent conditions (the shaded area shrinks).

Conceptually, however, developmental improvements in numerical judgement may also result from an improved ability to selectively at- tend to the representation of numerosity and amplify the contribution of numerosity to perceptual judgement while ignoring other quantita- tive information (average item size, density, total occupied area) that is also automatically extracted from sets of multiple items. According to this filtering hypothesis (see Fig. 1C), children get progressively better at teasing apart numerical from non-numerical quantitative variables when confronted with sets. Evidence suggests that already at an early age, children spontaneously estimate the variables of numerosity, size, and surface area (Cordes & Brannon, 2008, 2011). During development, the decision system would learn to focus on numerosity and to avoid interference from other continuous magnitudes, thus resulting in an increasingly accurate judgment. The existence of a congruity effects in numerical processing fits squarely with the filtering hypothesis. When asked to choose the numerically larger of two sets, human adults are less accurate when the size of the items, or the inter-item distance is incongruent with number, than when it is congruent (Gebuis & Reynvoet, 2012). Congruity effects are thought to arise from the fact that numerical and non-numerical dimensions are encoded in over- lapping sectors of parietal cortex, and in some cases, by the very same neurons (Harvey, Fracasso, Petridou, & Dumoulin, 2015; Pinel, Dehaene, Riviere, & LeBihan, 2001; Tudusciuc & Nieder, 2009). Be- cause of this overlap, brain areas downstream of those representing numerical and non-numerical dimensions may be confronted with the same problem that confronts multivariate classifiers, namely the iden- tification of relevant dimensions in a highly multidimensional set of neuronal responses (King & Dehaene, 2014, boX2).

Sharpening and filtering are not necessarily mutually exclusive learning mechanisms: both may jointly occur during development/education. However, they are qualitatively different. The former affects the precision of the representation (see Fig. 1B), while the latter affects the effectiveness of the decision system at discarding task-irrelevant representations (see Fig. 1C). Indeed, the two hypotheses make rather different predictions of the developmental time course of performance. As illustrated in Fig. 1B, if sharpening is the only mechanism, then there should be an overall reduction in error rates, particularly for stimuli close to the decision boundary, but not necessarily a reduction in the congruity effect: decreasing the noise without changing the decision boundary should result in increases in accuracy in both trials where number is incongruent with non-numerical dimensions and trials where number and the non-numerical dimensions are congruent. If only fil- tering is at work (see Fig. 1C), on the other hand, learning should dif- ferentially affect the congruent and incongruent trials: progress should be mostly observed on incongruent trials, but it should be absent on congruent trials. If there is only filtering, it is even possible that, in the course of learning, children would perform increasingly worse on con- gruent trials, as they would lose the benefit of a reliance on correlated helping variables. Such a behavior would clearly speak against the sharpening model, which would be unable to accommodate a decrease in performance. A third possibility is that, because sharpening and fil- tering are not mutually exclusive, they both occur during development: this would result in improvements occurring in both congruent and incongruent conditions, but more so in the incongruent conditions.

To test those predictions, we re-analyzed a large set of previously published psychophysical data where subjects of different ages and levels of numeracy were engaged in a common numerosity comparison task. Contrary to most previous research

(Bugden & Ansari, 2015; Gilmore et al., 2013; Szucs, Nobes, Devine, Gabriel, & Gebuis, 2013), here we varied the degree of congruity between numerical and non-numerical variables. We could therefore estimate the impact of nu- merical and non-numerical variables on the subjects' trial-by-trial choices, and examine their variations as a function of age and educa- tion, with the ultimate goal of directly contrasting the prediction of the sharpening and filtering models. We used two complementary approaches: the first, model based, used logistic regression, and the second, model free, used Shannon information. Finally, we presented a minimal simulation of learning in this task, based on the filtering hy- pothesis, and showing that a support vector machine, trained with dot stimuli, progressively learns to rely only the relevant numerical di- mension of the stimuli and mimics the observed human data.

2. Methods

2.1. Participants

Numerosity comparison data was obtained from a large group of subjects of different ages, education and cultures: 44 Italian kinder- garteners (age range = 3.6–6.2, mean = 5.1); 29 Italian school-aged children (age range 8–12, mean = 9.9); 20 Italian educated adults (age range = 22–33, mean = 26.6); 25 Italian dyscalculic children (age range = 8–12, mean = 10.3), and 38 Mundurucú children and adults of different education level (age range = 3–63, mean = 24.6). These data were previously reported only from the point of view of a change in Weber fraction (Piazza et al., 2010, 2013). Furthermore, in those publications we only included in the final analyses data from partici- pants who showed evidence of performing the task by consistently at- tending to number (as identified through a consistent effect of the nu- merical logratio on performance, with a minimum partial least square fit of the psychometric function = 0.2). Here, because our aim was to evaluate how performance evolves, we included the data of all parti- cipants. However, the results remained substantially unchanged when we restricted the present analyses to subjects included in the afore- mentioned publications.

2.2. Stimuli space

The stimuli consisted of pairs of arrays of black dots displayed within two white discs on either side of a central white fiXation point (see Fig. 2A for a couple of exemplar pairs of stimuli). On each trial, one of the two arrays contained either 16 or 32 dots (the reference, here- after referred to as n1). The paired array (the target, hereafter referred to as n2) contained between 10 and 22 dots (along the following 10- level continuum: 10, 12, 13, 14, 15, 17, 18, 19, 20, or 22 dots) when n1 was 16 and double those quantities when n1 was 32. For the pre- schoolers, about half of dyscalculics, and the adult participants, the trials with the most extreme n2 values (10 and 22 for n1 = 16, and 20 and 44 for n1 = 32) were omitted. In both versions of the experiment, perceptual variables were assigned to the stimuli such that, on half the trials the size of the dots and the average dot-to-dot distance were held constant across numerosities in the n2 arrays (as a consequence of this, numerosity of the n2 arrays was correlated with both the total area occupied by the items, and with the external envelope of the set). The n1 arrays paired with those stimuli were constructed such that these parameters varied simultaneously, and were randomly assigned to the n1 sets such that, from trial to trial, they covered all values assigned to the different n2 arrays. In this set of stimuli pairs, in one half the cu- mulative area occupied by the dots and the external envelope were congruent with number, and in the other half they were incongruent. On the other half of the trials, on the contrary, it was the area occupied by the items and the external envelope of the set that were held con- stant in the n2 arrays (thus numerosity of the n2 arrays was anti-cor- related with the individual dot size and the dot-to-dot distance), and varied randomly in the n1 arrays. In this set of stimuli pairs, half had the size of individual dots and the external envelope of the set which were congruent with number and the other half they were incongruent (see Dehaene, Izard, & Piazza, 2005, for the stimulus generation pro- gram). This design ensured a large variability in continuous features across trials, such that numerosity judgments could not be above chance if subjects attended solely to one of these non-numerical para- meters throughout the experiment. Stimuli remained on-screen until participants gave their response, which consisted of pressing the button on the computer keyboard that corresponded to the position of the more numerous set. Subjects were explicitly instructed to respond as quickly as possible without counting. Previously reported response times (in the range of 1–2 s) indicated that they were not counting, but were using an estimation strategy.

For the sake of coherence with the existing literature, we here adopt the terminology of DeWind, Adams, Platt, & Brannon (2015), who used a parametric modeling approach similar to ours, also recently with children (Starr, DeWind, & Brannon, 2017). We therefore defined our stimuli space using four variables (all expressed in terms of number of piXels), which, combined in pairs, univocally define number: item surface area (the area occupied by a single dot, hereafter ISA) and total surface area (the item surface areas multiplied by the number of items, hereafter TSA); field area (also sometimes referred to as "convex hull", indicating the portion of the space where dots actually fall into, hereafter FA) and sparsity (the field area divided by the number of items, hereafter Spar). Because these 4 variables are clearly inter-de- pendent, and because their full combination is fully confounded with number, it was not possible to simultaneously test the effect of all 4 together with that of number. The two summary parameters "Size" and "Spacing" defined in the original paper by DeWind were not orthogonal with number, as our stimuli were not constructed to satisfy this prin- ciple. Thus, in order to estimate the combined impact on the non-nu- merical variables on subjects' decision without arbitrarily selecting a subset of such variables, first we computed the component of each variable that was orthogonal to number, obtaining 4 new non-numer- ical variables. To do this we subtracted from each variable its scalar product with number (i.e. the component that was parallel to number). Moreover, because the scope of the current study was not to precisely determine which specific physical characteristic of the stimuli has the strongest influence on numerical judgment, but rather to pit number vs. the non-numerical quantitative dimensions considered as a whole, we decided to condense the non-numerical variables in a single summary measure, called "Non-Numerical Dimension" (NND) which we defined as the first principal component (estimated by means of Principal Component Analysis) of the set of the 4 previously defined non-nu- merical variables. This component explained 98.9% of the variance of the 4 different non-numerical parameters, proving that it was a good summary measure. The relative weights of the 4 non-numerical di- mensions used by PCA to build the NND were 0.557, 0.487, 0.473, 0.467 for sparsity, ISA, TSA and FA, respectively, indicating that all equally loaded onto the NND. This new variable was, by construction orthogonal to number, thus allowing us to evaluate its impact on sub- jects' choice independently from that of number (see Fig. 2A for ex- amples of pairs of stimuli where number and the NND were congruent (left) vs. incongruent (right)). All analyses were also replicated separately for each non-numerical variable, and the results were essentially identical.

3. Analyses and results

We started by separately analyzing, for each group, the trials in which the non-numerical dimension was congruent with the numerical dimension, and those for which it was incongruent. Incongruent trials represented about 50% of trials (49% for adults, 50% for Mundurucús, 50% for preschoolers, 51% for school kids, and 49% for dyscalculics).
Overall accuracy was higher on congruent than on incongruent trials in all groups but for the 8–12 non-dyscalculic children (an ANOVA Group by Congruency yielded a significant effect of Group $F(2, 180) = 26.79$, $p < 0.001$, $\eta 2 = 0.23$, of Congruency $F(1, 180) = 176.96$ $p < 0.001$,

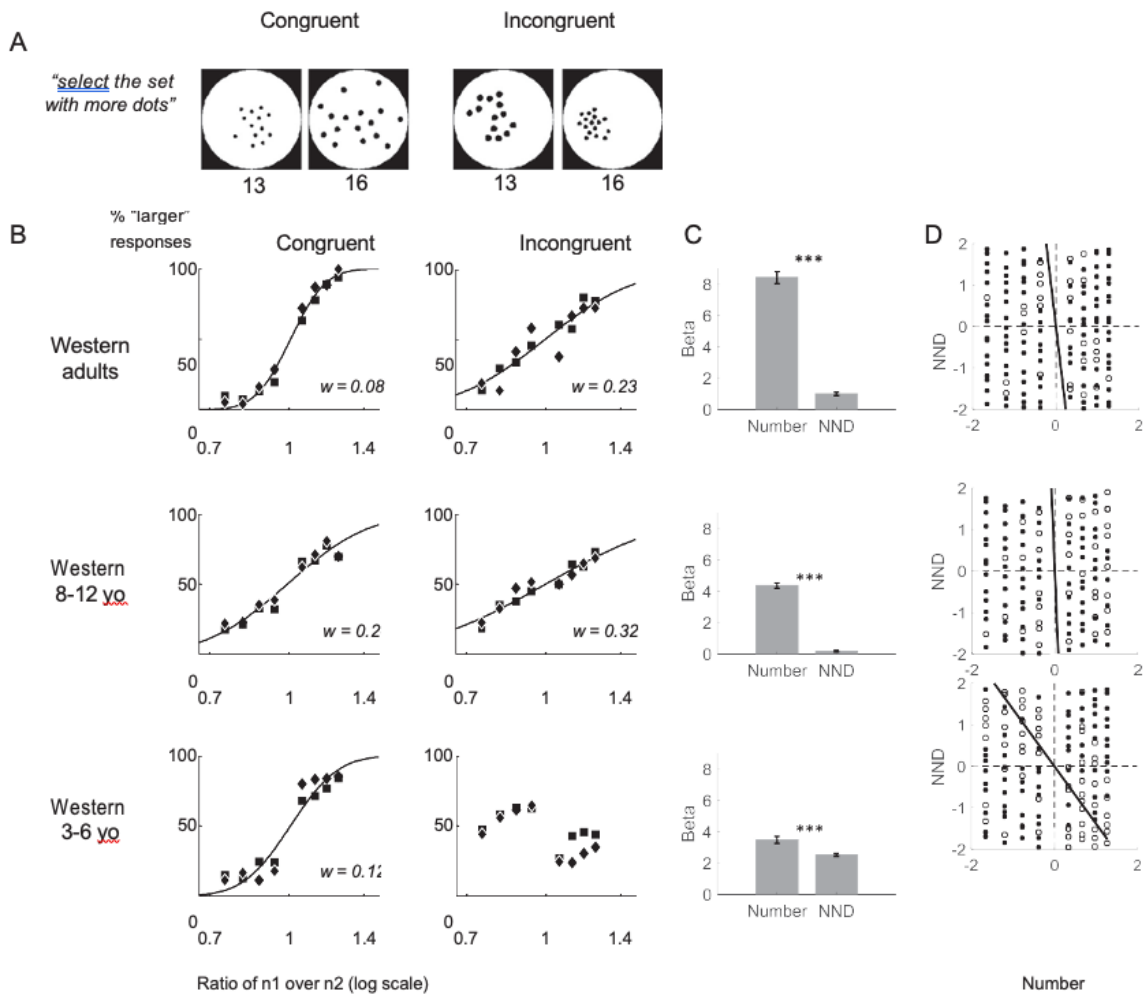


Fig. 2. Influence of congruity between number and the non-numerical dimensions on the number comparison task for the three groups of western non-dyscalculic subjects. (A) EXamples of stimuli where number is either congruent or incongruent with other non-numerical dimensions. (B) Psychometric functions of three age groups: pre-schoolers (Western 3–6 yo), school children (Western 8–12 yo) and adults (Western adults). Each plot represents the percentage of responses in which the comparison array (n2) was reported as larger than the reference array (n1), as a function of the log n2/n1 ratio (black squares: n1 = 16; black diamonds: n1 = 32). The curves are normal cumulative-distribution fitting functions used to estimate the internal weber fractions. Left, trials where number and the non-numerical dimension were congruent. Right, incongruent trials. Psychometric curves could be fit for all groups and conditions apart from the trials in incongruent conditions in the kindergarten group. (C) Beta values from the multiple logistic regression computed at the group level with two predictors (number and the specified non- numerical dimension (NND); error bars are SEM of the estimated beta). (D) Decision boundary estimates projected on the real stimuli space: dots represent a sub set of real stimuli (for clarity reason, chosen by random down-sampling the full set) as the unique combination of number and NND (expressed here as the SD normalized logratio of number and NND). Empty dots represent error trials, while filled dots correspond to correct trials. The solid black line represents the decision boundary (see text). The systematic presence of errors in the top left and bottom right quadrant (incongruent conditions) in preschoolers indicates that they used a sub-optimal decision boundary, due to interference from the irrelevant dimension.

$\eta 2 = 0.50$ and critically of their interaction $F(2, 180) = 40.13$, $p < 0.001$, $\eta 2 = 0.31$; planned comparisons indicated significantly more errors in the incongruent vs congruent for all groups but the 10 years old school kids (Adults = 64% vs 87%; Preschooler = 36.4% vs 82.3%; 10 years old school kids = 66.1% vs 74.6%). The absence of a congruity effect in 10 years old school kids originated from the fact that, when compared to preschoolers, they improved in the incongruent trials (from 36.4% ± 1.8% to 66.1% ± 3.1% accuracy, $t(71) = 8.86$, $p < 0.001$), but they actually got worse on congruent trials (from 82.3% ± 1.6% to 74.6 ± 3.0% accuracy, $t(71) = 2.46$, $p = 0.017$); compare second and third rows in Fig. 2B. This is precisely what is predicted by the filtering hypothesis. Average performances for congruent and incongruent trials in each group are represented in Fig. 2B and fitted by psychometric curves which we computed following our our previously published methodology (Piazza et al., 2010, 2013). The psychometric curves were, for all groups, steeper in the congruent than in the incongruent condition, yielding different estimates of the internal weber fraction

3.1 Logistic regression approach

To investigate the relative weight of numerical and non-numerical dimensions, we used a multiple logistic regression analysis on the trial- by-trial performance to estimate the contributions of two parameters, the logratio of the numerosities of the two sets and of their non-nu- merical dimensions. This approach was first used at the group level, by pooling the responses of all trials and of all subjects within each group. In all groups, the beta coefficients of both predictors were highly sig-
nificant (all $p < 0.001$) and all models fitted the data well (Hosmer–Lemeshow test Goodness of Fit $p > 0.05$), except than for the preschoolers group (HL GOF $p = 0.007$). In all groups the beta coeffi- cient of number was significantly higher than that of the non-numerical dimension (all p values < 0.001) (adults: $z = 18.54$; school kids: $z = 23.76$; kindergartener: $z = 4.01$; dyscalculics: $z = 12.99$; Mundurucús: $z = 22.94$), see Figs. 2C and 3B. Because the ratio be- tween the estimated beta coefficients of number and of the NND can be readily interpreted as the slope of the decision boundary on the stimuli

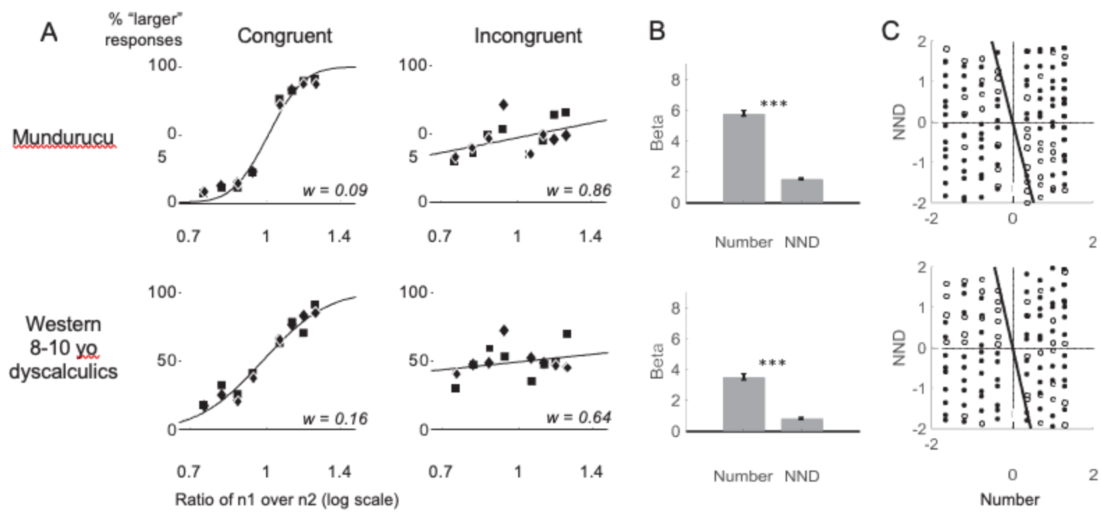


Fig. 3. Influence of congruity between number and the non-numerical dimensions on the number comparison task in Mundurucús and dyscalculic subjects. (A) Psychometric functions of two groups. Each plot represents the percentage of responses in which the comparison array (n2) was reported as larger than the reference array (n1), as a function of the log n1/n2 ratio (black squares: n1 = 16; black diamonds: n1 = 32). The curves are normal cumulative-distribution fitting functions used to estimate the internal weber fractions. Left, trials where number and the non-numerical dimension were congruent. Right, incongruent trials. (B) Beta values from the multiple logistic regression computed at the group level with two predictors (number and the specified non-numerical dimension (NND); error bars are SEM of the estimated beta). (C) Decision boundary estimates projected on the real stimuli space. Dots represent stimuli as the unique combination of number and NND
(expressed here as the SD normalized logratio of number and NND). Empty dots represent error trials, while black dots correspond to correct trials. The solid black line represents the decision boundary (see Section 2). For clarity reasons, the dots represent only a subset of all the real stimuli (chosen by random sub-sampling of the full set).

space (a graphical representation of this can be found in Fig. 1C), we computed this ratio for each group and transformed it into an angle (defined as the inverse tangent of the slope). In Fig. 2D and Fig. 3C we plot the resulting decision boundary for each of the five groups. The angle was tilted away from the optimal decision boundary slope ($\vartheta = 90°$) in kindergarteners, and with age it approached the optimal 90° inclination (adults: $\vartheta = 83.28 \pm 0.71°$; school kids: $\vartheta = 87.66 \pm 0.68°$; kindergartener: $\vartheta = 54.00 \pm 2.72°$). The tilting of the decision boundary during development/education is in line with the predictions of the filtering hypothesis (compare the bottom and top panels in Fig. 2D with right and left panels in Fig. 1C). We also explored these effects in dyscalculic children, in whom we had previously re- ported a deficit of the Weber fraction (Piazza et al., 2010). The average inclination of the decision boundary for the dyscalculics was $\vartheta = 77.56 \pm 1.19°$, significantly smaller than the value of $\vartheta = 87.66 \pm 0.68°$ observed in their age – and IQ matched controls ($z = 7.38$, $p < 0.001$), see Figs. 2D and 3C. In the Mundurucús, the average boundary angle was $\vartheta = 75.49 \pm 0.59°$.
In order to further support these group results with statistical analyses that take into account the inter-individual differences, we ran the same logistic regression analysis at the single subject level. Results confirmed that the beta coefficient of number was significantly higher than that of the non-numerical dimension for all groups but the kin- dergarteners

(adults: $t(19) = 11.4$; school kids $t(28) = 10.8$; dyscalcu- lics $t(24) = 4.93$; Mundurucús $t(37) = 7.59$; all $ps < 0.001$; kinder- garteners $t(43) = 0.11$, $p = 0.91$). This finding indicates that even when a strong influence of the physical parameters of the stimuli was present, number remained the dominant variable driving subjects' judgements for all groups but the younger ones (see Figs. 2C and 3B). With these single-subject measures in hand, we next explored the developmental trajectory of the two beta values (for number and the NND) as a function of age and education (Fig. 4). In the Western edu- cated non-dyscalculic subjects, age led to an increase in the impact of number (linear regression of age on the estimated beta coefficient of number in the previously described logistic regression $r2 = 0.27$, $p < 0.001$) and a decrease in the impact of the irrelevant non-nu- merical dimension ($r2 = 0.11$, $p = 0.01$), see Fig. 4A. These effects, however, are mostly seen when considering all age groups together: when looking separately within each group (kindergarteners, school kids, and adults), significant effects of age are seen only within the kindergarteners (regression between age and the beta estimate of number and of the NND, both $p < 0.05$). This is potentially due to the smaller sample sizes within the two older groups. Alternatively, it could indicate that our effects display true discontinuities during develop- ment. Testing this possibility would require a constant number of subjects for each age as well as a continuous sampling of age during the life-span. Future studies could address this question. In our sample, both within and across each group, age and education are highly cor- related. Thus, to specifically test the effect of education, we moved to the Mundurucús, for whom education is limited and largely in- dependent of age. First, we tested a linear regression of education on the two coefficients in the entire population of Mundurucús tested. While the effect of number showed no increase with education ($p = 0.98$) (see Fig. 4B, left panel), we observed that education induced a dramatic decrease in the effect of the non-numerical dimension, even when regressing out age (regression of education on the age standar- dized residuals $r2 = 0.32$, $p < 0.001$) (see Fig. 4B, right panel).

To further analyze this effect, we focused on the adult Mundurucús, in whom age and education were best dissociated. The average effect of the non-numerical dimension for uneducated adult Mundurucús (N = 7) was 8.28, much larger than the value of 1.1 observed in edu- cated Italian adults, $t(25) = 7.68$, $p < 0.01$. This value dropped to 2.02 in adult Mundurucús who went to school for at least one year (N = 13). Thus, there was a highly significant difference in the amount of non-numerical interference between educated and uneducated Mundurucús, t(18) = 4.6 , p < 0.001 (see Fig. 3), even though the two groups did not differ in age (t(18) = −1.32, p = 0.20), or in overall response times (t(18) = 1.11, p = 0.28).

To summarize the results of the multiple regression analyses so far, all groups, including adults, suffered from interference from irrelevant non-numerical dimensions, and the modulation of this interference was the dominant factor in the developmental change in performance with age, education or in dyscalculic subjects. Those findings appear highly compatible with the predictions of the filtering hypothesis.

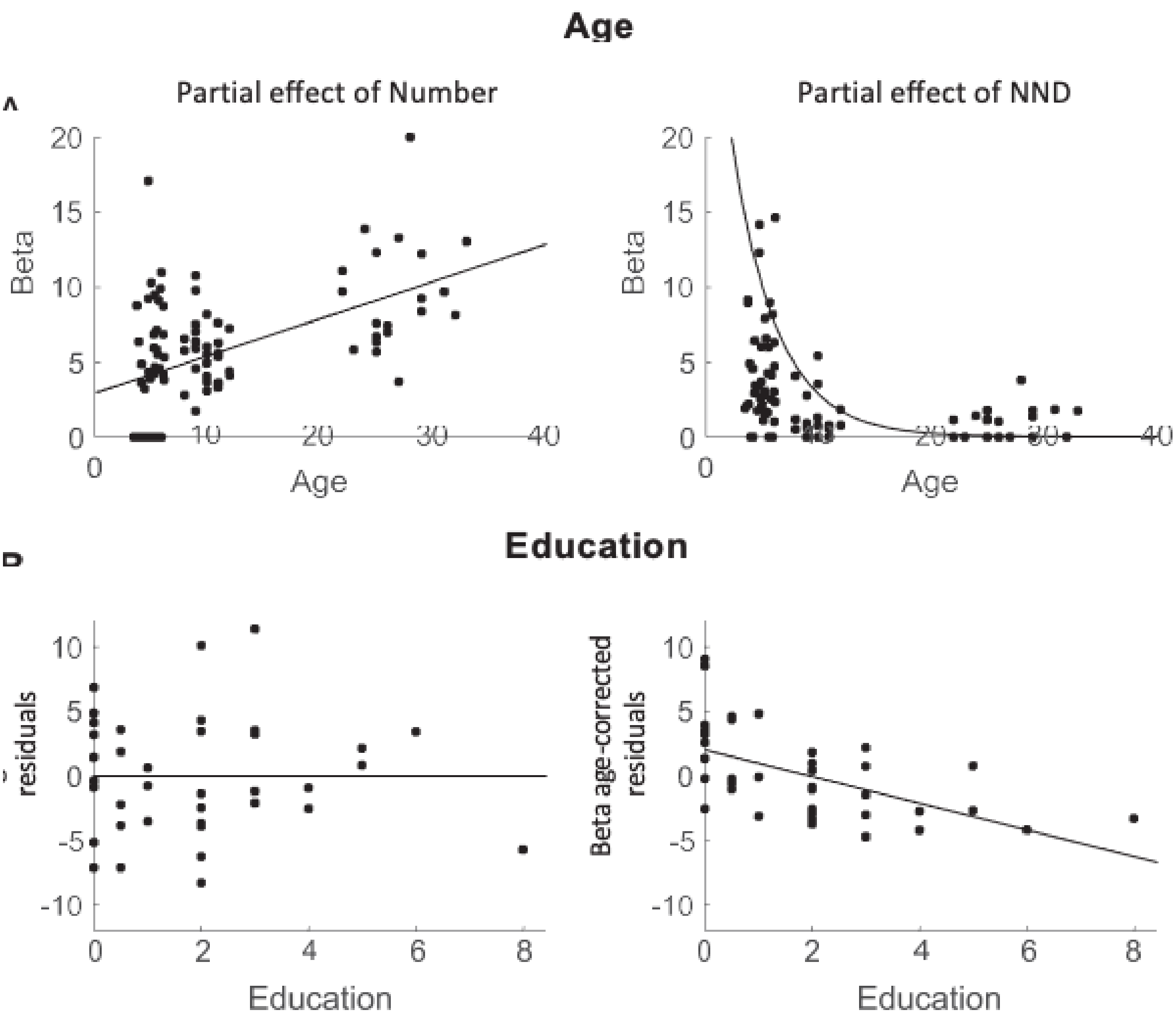


Fig. 4. Single-subject analyses of the effects of age and education. (A) Single subject estimates of the beta values from the multiple logistic regression, plotted as a function of age in the typically devel- oping population together with the best-fitting curve. (B) Beta values from the multiple logistic regression corrected

for age, plotted as a function of level of education in the Mundurucú population together with the best-fitting curve.

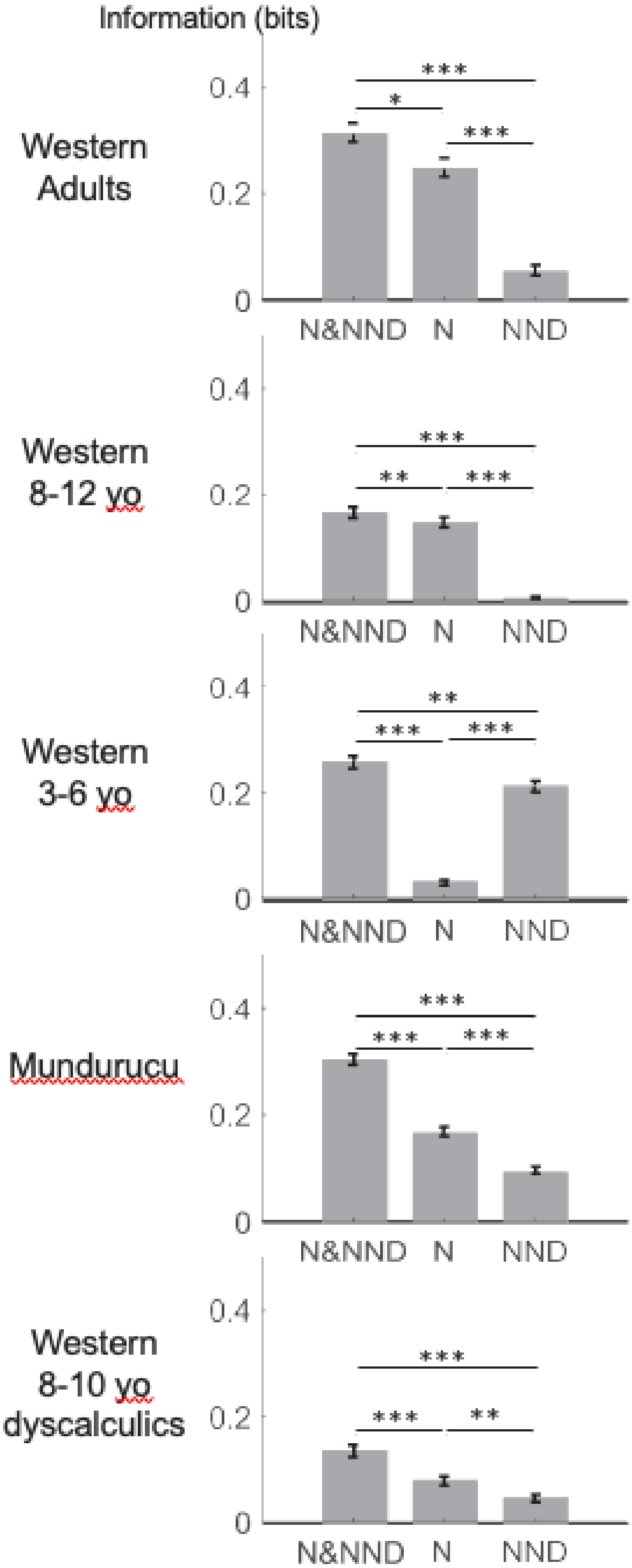
Because the regression analyses performed so far assume linear effects and are based on a specific model of binary choice, which was unable to capture the data from kindergarteners, we complemented and strengthened these initial analyses by using mutual information (Shannon, 1948), a general and assumption-free approach which we used to estimate the effect of number and the NND on subjects' choice. Unlike correlation measures such as logistic regression, mutual in- formation captures all possible statistical relationships among vari- ables, including linear and nonlinear ones (Quian Quiroga & Panzeri, 2009; Shannon, 1948). Mutual information, I(C;S), carried by the considered dimension of the stimulus s (s = ratio of the two numer- osities and/or ratio of the two NNDs) belonging to set S to the subject's choice c (belonging to set of possible choices C = {larger, smaller}), is defined as follows:

$$I(C;S) = \sum_{s \in S} \sum_{c \in C} P(s,c) \log_2 \frac{P(s,c)}{P(s)P(c)} \quad (1)$$

where P(s,c) is the probability that in a given trial the subject chose c and stimulus with feature s was presented, P(c) is the probability of choice c unconditional on the stimulus, and P(s) is the probability of presentation of stimulus with feature s. I(C;S) quantifies, in units of bits, the average reduction of uncertainty about which choice was taken based on a single-trial observation of the numerical (and/or non-numerical) features of the stimulus. I(C;S) is zero bits only when the choice is fully independent from the features of the stimulus, as in that case no knowledge about choice can be gained by observing the features of the stimulus. I(C;S) reaches its maximal value (1 bit if stimuli and choices are equiprobable) if choice depends faithfully on the stimulus (i.e. when the subject always performs the task correctly). Unlike simpler trial-averaged measures of task performance, such as average percent correct, I(C;S) takes fully into account the distribution of correct choices and errors in evaluating the relationships between choice and specific stimulus variables. Unlike the beta coefficient of a multiple regression, I(C;S) expresses the results on a physical scale of bits that can be directly interpreted in terms of reduction of uncertainty. Information in Eq. (1) was computed according to established procedures (Panzeri, Senatore, Montemurro, & Petersen, 2007), by first discretizing stimulus values, then plugging the empirical stimulus-choice probabilities into Eq. (1) and finally using the Panzeri-Treves bias correction to remove the limited sampling bias (Treves & Panzeri, 1995). We first computed the mutual information between number and the NND together and the subjects' choice, and then between either variable and choice separately (see Fig. 5), on the pooled data for each group.

Results, plotted in Fig. 5, revealed, for all groups, three important results. First, the information carried to choice was significant, in all groups, for both number and the NND (number p values < 0.001 for all groups; NND p values < 0.001 for adults and kidergarteners, $p = 0.008$ for school kids). This means that both stimulus variables, in all groups, influenced response choice. Second, we found (Fig. 5) that, for all groups, the information that number and NND jointly carried to choice was larger than the information that either stimulus variable alone carried to choice (Comparison between information carried jointly and information carried by number: $I(C;N,NND) = 0.315 \pm 0.018$ bits, $I(C;N) = 0.249 \pm 0.018$ bits, $z = 2.62$ and $p = 0.004$ for adults; $I(C;N,NND) = 0.165 \pm 0.010$ bits, $I(C;N) = 0.147 \pm 0.010$ bits, $z = 1.28$ and $p = 0.010$ for school kids; $I(C;N,NND) = 0.260 \pm 0.012$ bits, $I(C;N) = 0.033 \pm 0.005$ bits, $z = 17.67$ and $p < 0.001$ for kindergarteners. Comparison between information carried jointly and information carried by NND: $I(C;N,NND) = 0.315 \pm 0.018$ bits, $I(C;NND) = 0.056 \pm 0.010$ bits, $z = 12.82$ and $p < 0.001$ for adults; $I(C;N,NND) = 0.165 \pm 0.010$ bits, $I(C;NND) = 0.005 \pm 0.002$ bits, $z = 15.00$ and $p < 0.001$ for school kids; ; $I(C;N,NND) = 0.260 \pm 0.012$ bits, $I(C;NND) = 0.214 \pm 0.010$ bits, $z = 2.88$ and $p = 0.002$ for kindergarteners). This indicates that neither number nor the NND alone is able to determine subjects' choice with the same accuracy as the combination of the two.

Moreover, results in Fig. 5 show also that number and NND carried approximately independent information to the choice, because the choice information present when considering jointly both variables



Information (bits)
Western
Adults
0.4
0.2
0
***
*
***
N&NND
N
NND
Western
8-12 yo
0.4
0.2
0
***
**
***
N&NND
N
NND
Western
3-6 yo
0.4
0.2
0
**
***
***
N&NND
N
NND
Mundurucu
0.4
0.2
0
***
***
***
N&NND
N
NND
Western
8-10 yo
dyscalculics
0.4
0.2
0
***
***
**
N&NND
N
NND

Fig. 5. Information (expressed in bits) carried by the two stimulus features (number (N) and the non-numerical dimension (NND)) to choice, considered together (left column of the plot) and separately (central column and right column for N and the NND, respectively). Information was computed at the group level, pooling together all trials for subjects in each group (error bars are SEM of the estimated information).

approXimately equals the sum of the choice information carried by the two variables separately. This means that number and NND carried complementary (non-redundant) information to the choice.

Finally, by contrasting the amount of information carried to choice by number and the NND, we found that for adults, the information carried by number (I(C;N) = 0.249 ± 0.018 bits) was significantly higher than the one carried by the non-numerical dimension (I (C;NND) = 0.056 ± 0.010 bits) ($z = 9.70$, $p < 0.001$). For the schoolkids we had a similar pattern (I(C;N) = 0.147 ± 0.010 bits, I (C;NND) = 0.005 ± 0.002 bits, $z = 14.45$, $p < 0.001$). On the contrary, for the kindergarteners the opposite pattern was observed (I (C;N) = 0.033 ± 0.005 bits, I(C;NND) = 0.214 ± 0.010 bits, $z = 15.80$, $p < 0.001$) (see Fig. 5). Thus, the information analysis provided evidence that, unlike the other groups of western non-dyscalculic subjects, preschoolers mainly based their decisions on the NND, and much less on number.

We then analyzed dyscalculics and Mundurucús and found that the choice information carried by number was significantly higher than the choice information carried by NND (for dyscalculics: I (C;N) = 0.080 ± 0.008 bits, I(C;NND) = 0.048 ± 0.007 bits, $z = 2.97$, $p = 0.001$; for Mundurucús: I(C;N) = 0.166 ± 0.009 bits, I (C;NND) = 0.094 ± 0.007 bits, $z = 6.62$, $p < 0.001$). However, dyscalculics differed importantly from their age and IQ matched control group: information from the NND was higher (0.048 ± 0.007 bits versus 0.005 ± 0.002 bits, $z = 5.91$, $p < 0.001$), while that from number was lower (0.080 ± 0.008 bits versus 0.147 ± 0.010 bits, $z = 5.21$, $p < 0.001$), $p = 0.01$), see Fig. 5.

To further investigate the effect of choice information taking into account single subjects' variability we re-computed information at the single subject level. First, we confirm that for all groups but the kindergarteners mainly based their decisions on the basis of number (for adults: I (C;N) = 0.302 ± 0.022 bits, I(C;NND) = 0.059 ± 0.012 bits, t (19) = 9.75, $p < 0.001$; for school kids I(C;N) = 0.187 ± 0.017 bits, I (C;NND) = 0.089 ± 0.019 bits, t(28) = 3.27, p = 0.003; for the kinder-garteners: I(C;N) = 0.097 ± 0.008 bits, I(C;NND) = 0.251 ± 0.025 bits, t(43) = 5.60), $p < 0.001$).

Then, similarly to what we did in our previous logistic regression analyses (Fig. 4), we investigated whether age and education modulated information from number and the NND by performing regression analyses (see Fig. 6). In Western educated non-dyscalculic subjects (Fig. 6A), age led to an increase in the information carried by number (r2 = 0.47, $p < 0.001$) and a decrease in that carried by the non-numerical dimension (r2 = 0.19, p = 0.001). Similarly as with the betas of the multiple regressions, here the age effect failed to reach significance if we restricted the regression analysis within each age group, potentially due to small sample sizes within each group. Again, in order to tease apart the effect of age and that of formal education we moved to the analyses of the Mundurucús. Results confirmed that education was the principal factor underlying these developmental changes (see Fig. 6B): more educated Mundurucús increasingly based their decisions on number (r2 = 0.22, p = 0.005), and displayed less interference from non-numerical information (r2 = 0.30, $p < 0.001$).

Taken together, the mutual information results confirm that while all groups, including adults, base their decision at least in part upon task-irrelevant non-numerical dimensions, the reduction of this inter- ference is the dominant factor in the developmental change in perfor- mance with age and education.

To better differentiate between the predictions of the filtering and of the sharpening hypotheses, we finally used information theory to in- vestigate the developmental/educational trajectory of the congruity effect. In this analysis, for each subject we separated congruent and incongruent trials. While the filtering hypothesis predicts that the general improvement due to age and education is due to a selective increase of information on number in incongruent trials, the sharpening hypothesis predicts that information on number should increase in both congruent and incongruent trials. In order to maximize the congruency effect, we selected the trials where the amount of congruency was very high and very low, defined as those for which number and the NND were either larger than the 70th or smaller than the 30th percentile of each distribution (congruent trials = 23.3% of the total number of trials, incongruent = 23.1%). On this subset of trials, we computed, for each subject, the information carried to choice by number. In Fig. 7A we present the developmental trajectory of information carried by number separately for congruent and incongruent trials. In the upper part of each plot, where the axis goes from 0 to 1, bottom to top, 1 represents the cases where information is totally coherent with number: a subject who scores 1 is a subject whose choice was always coherent with number and thus who was always correct in the number com- parison task. In the bottom part of the figure, where the axis goes from 0 to 1 in the opposite direction, 1 represents the cases where in- formation is totally incoherent with number: a subject who scores 1 here is a subject whose choice was always incoherent with number and thus always wrong. Zero represents chance with respect to the task. The results indicate that while in congruent trials the information on number remains constant during development (linear regression between information and age $r2 = 0.0005$, $p = 0.8$), it dramatically changes in the case of incongruent trials: while preschoolers' choices

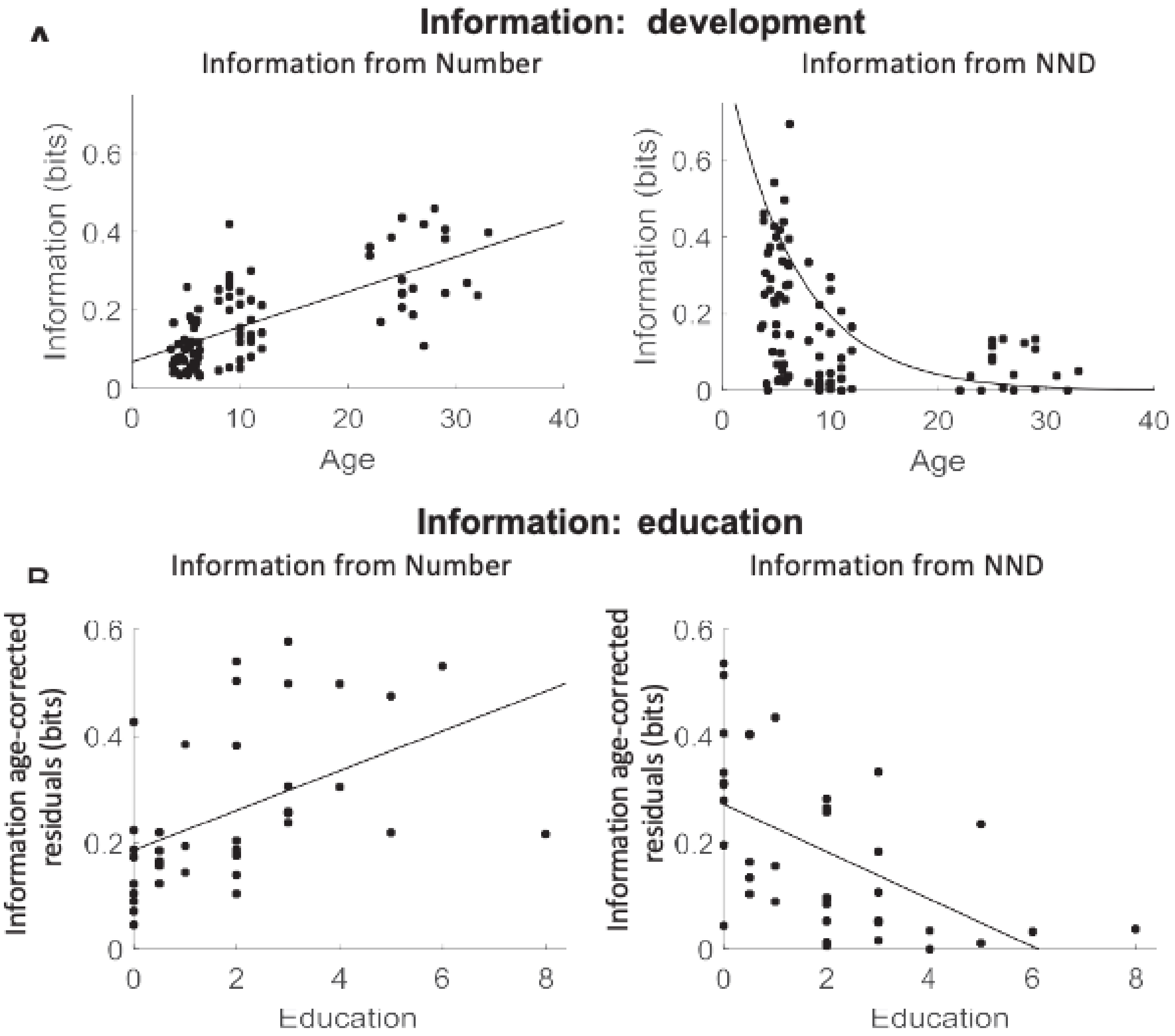


Fig. 6. (A) Single-subject analyses of the effect of age on information from number and the NND. Single subject estimates of the impact of number and of the NND on decision in the left and right panel, respectively. Data are plotted as a function of age in the typically developing population to- gether with the best-fitting curve. (B) Single-subject analyses of the Mundurucú data, investigating the effect of education on information from number and the NND. Single subject estimates of the impact of number and of the NND on decision in the left and right panel, respectively. Data are plotted as a function of age in the typically developing popu- lation together with the best-fitting curve.

were systematically consistent with the NND and inconsistent with number, with age and education preschoolers became able to filter out the interference from the NND thus responding mainly according to number (linear regression between age and information, considering the coherent and incoherent information as a continuum from totally incoherent to totally coherent, r2 = 0.14, p < 0.001).

As for the previous analyses, the Mundurucús data were used to separate developmental maturation from education. The results,

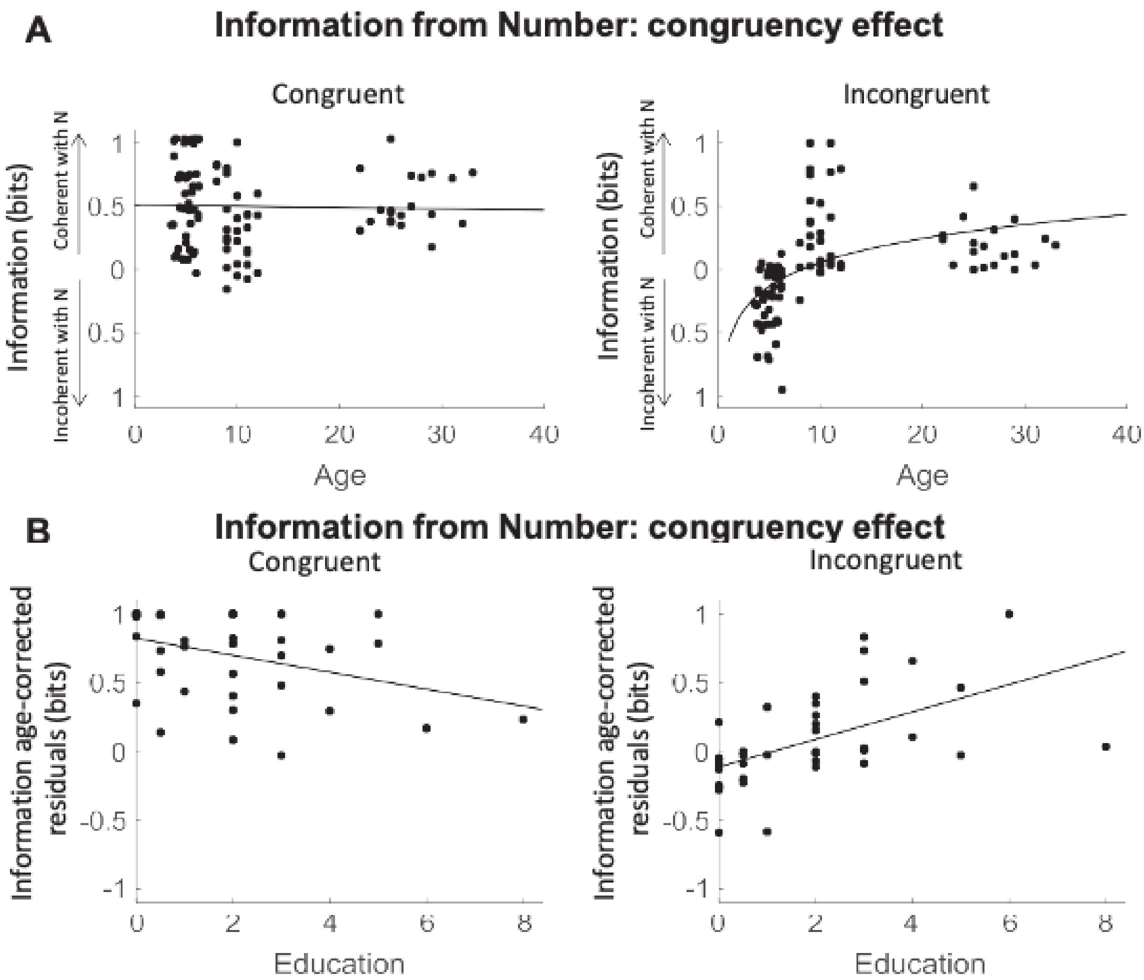


Fig. 7. (A) Single subject estimates of the impact of number in the strongly congruent and strongly incongruent trials in the left and right panel, re- spectively. Data are plotted as a function of age in the typically developing population together with the best-fitting curve. (B) Age corrected estimates of the impact of number in the

strongly congruent and strongly incongruent trials in the left and right panel, respectively. Data are plotted as a function of education in the Mundurucú population together with the best-fitting curve.

depicted in Fig. 7B, indicate that education (once correcting for age) is a factor that yields important changes: while in congruent trials it yields to a slight decrease in the information from number ($r2 = 0.13$, $p = 0.02$), it yields a dramatic increase in incongruent trials ($r2 = 0.30$, $p < 0.001$). This indicates that education acts by improving the ability to focus on number and filter out irrelevant non-numerical quantitative features of the sets. As a last step, as a proof-of-concept, we examined if the reported changes in performance as a function of development and education could be simulated using a minimal computational model. We trained a simple multivariate classifier (Cortes & Vapnik, 1995) on our task using a linear Support Vector Machine, as implemented in the basic SVM function in the Matlab software. On each trial, the classifier was given as input the 5 dimensions that characterized each pairs of sets (number, ISA, TSA, FA, and Spar), for a total of 10 features, and was trained to provide as output which of the two sets contained the larger number of dots. We set the initial parameters of the model (noise level = 10%, and proportion of congruent vs. incongruent trials = 75% potentially mimicking children's experience in the natural environment, where number and the physical variables often co-vary) such that during the initial phases of training the model approXimated the performance of our group of kindergarteners. Then we evaluated the evolution of the classifier performance using a typical 95%–5% cross-validation scheme (training the classifier on 95% of the data and testing on the remaining 5%), reporting average performance every 10 training trials. The re- sults, depicted in Fig. 8, indicate that the classifier initially performed quite poorly, and especially so on incongruent trials, but then quickly tuned in to the most relevant input dimension (number) as training progressed. Fig. 8C shows the performance of the classifier as a function of the log ratio of the compared numbers. During the initial stages of training, performance on congruent trials already exhibited a standard psychometric curve as a function of log ratio, while it remained flat on incongruent trials, mimicking the performance of preschoolers, dys- calculics, and uneducated adult Mundurucús (compare Fig. 8C, left plots, with Fig. 2B (preschoolers) and Fig. 3A (Mundurucú Indians and western dyscalculics)). After training, performance improved such that the derived psychometric curves of the model nicely reproduced per- formance observed in educated adults (compare Fig. 8C, right plots, with Fig. 2B (adults)). Thus, this implementation of the filtering idea captures the bulk of the data, although we emphasize that this model is only a proof-of-principle, and that more work will be needed to systematically explore the range of parameter values and to quantify the model fit.

4. Discussion

Using two complementary analytical approaches (regression and mutual information) on a large data set from human subjects of dif- ferent ages and levels of math education, we have demonstrated that the improvement in numerosity comparison performance with age and education mainly consists in an improved ability to focus on number and ignore other non-numerical dimensions of the sets (here called “filtering” mechanism). While the existence of congruity effects was previously reported (Gilmore et al., 2013; Szucs et al., 2013), here, by parametrically varying the degree of congruity between number and other variables, we go beyond a binary congruency approach. By comparing numerate and non-numerate children and adults, moreover, we dissociate education from maturation effects, and demonstrate that education is a key factor whose primary effect is to strongly decrease the impact of the non-numerical features on numerical decision making. We finally demonstrate that dyscalculic subjects’ impairment in numerosity comparison is mainly due to a disproportionally high interference effect (see (Bugden & Ansari, 2015) for a similar result), suggesting that dyscalculics profit less from formal education compared to their typically developing peers.

The bulk of these results fit squarely with the “filtering hypothesis”, that sees numerosity comparison as a multivariate classification pro- blem: during learning, decision units must determine which of their inputs provide cues to number, and which provide cues to non-numerical co-varying parameters, and they must learn to selectively at- tend to the first ones. The filtering hypothesis predicts that learning- related improvements should be mainly confined to trials where number information conflicts with information coming from continuous parameters of the sets (incongruent trials). Early math education, by highlighting the invariant nature of number with respect to the con- tinuous quantitative features of sets, is probably the best source of training for such a “mental classifier” tool.

The most stringent empirical evidence in favor of the filtering hypothesis stemmed from our analysis of information. We quantified how much information about number is carried to the subject’s choice, and how such information changes with age and education. Finally, we compared the developmental trajectory of information about number in congruent vs. incongruent trials. In congruent trials, information about number did not increase with age and education, whereas it did in- crease with age and education in incongruent trials. These observations, while fitting squarely with the predictions of the filtering hypothesis, are at odd with the sharpening hypothesis, which assumes that ma- turation and education mainly act by sharpening the internal re- presentation of numerosity, and thus predicts improvements in both congruent and incongruent trials.

The present research also speaks to the general field of perceptual learning and decision making, where it identifies the control of inter- ference as a key ingredient. In the past, scholars had interpreted im- provements in perceptual decision making as a sharpening of the re- presentational codes (Gilbert, Sigman, & Crist, 2001; Goldstone, 1998; Yang & Maunsell, 2004). Accordingly, the lifespan improvements in numerosity discrimination were thought to be due to a progressive sharpening of the tuning curves of parietal number-coding neurons (Piazza, Pinel, Le Bihan, & Dehaene, 2007; Verguts & Fias, 2004). However, representation sharpening is only one of the potential me- chanisms underlying perceptual learning. Another potential mechanism is the improved selection of task-relevant input from noisy sensory re- presentations (Dosher & Lu, 1998; Petrov, Dosher, & Lu, 2005). This idea was previously applied to the discrimination of simple object features, such as orientation or contrast, where the signal to filter out was background noise. Here we extend this view to number, which is a property of a set, and where the irrelevant signals that must be filtered out (density, average size, total occupied area) are other properties of the very same set, not simply background. According to the filtering hypothesis, learning is implemented through a mechanism akin to the training of a linear filter whose discrimination vector gets aligned to the dimension which is relevant to the task at hand, excluding the spurious correlations with other dimensions.

The filtering mechanism may take place in the prefrontal cortex, where neurons flexibly encode the relevant features of complex, mul- tidimensional stimuli according to context and task relevance. Support for this idea comes from a study of perceptual decision making in monkeys (Mante, Sussillo, Shenoy, & Newsome, 2013): during a color or motion discrimination task on colored moving stimuli, prefrontal cortex neurons initially represented both the color and motion dimen- sions, but during decision-making, only the task-relevant dimension got projected to the choice dimension. In the number domain, this hy- pothesis was recently evaluated in macaques (Viswanathan & Nieder, 2015): neurons in both parietal and prefrontal cortex were recorded during a numerosity discrimination training, where the number and color of stimuli varied from trial to trial. Training did not change the tuning curves of number neurons (thus, the sharpening hypothesis was rejected in this region). However, in prefrontal cortex only, training increased the number of number-selective neurons and their degree of selectivity for number vs. color (but again not the width of their tuning curves). These observations suggest that learning in monkeys was not dependent upon a sharpening of the primary parietal representation of

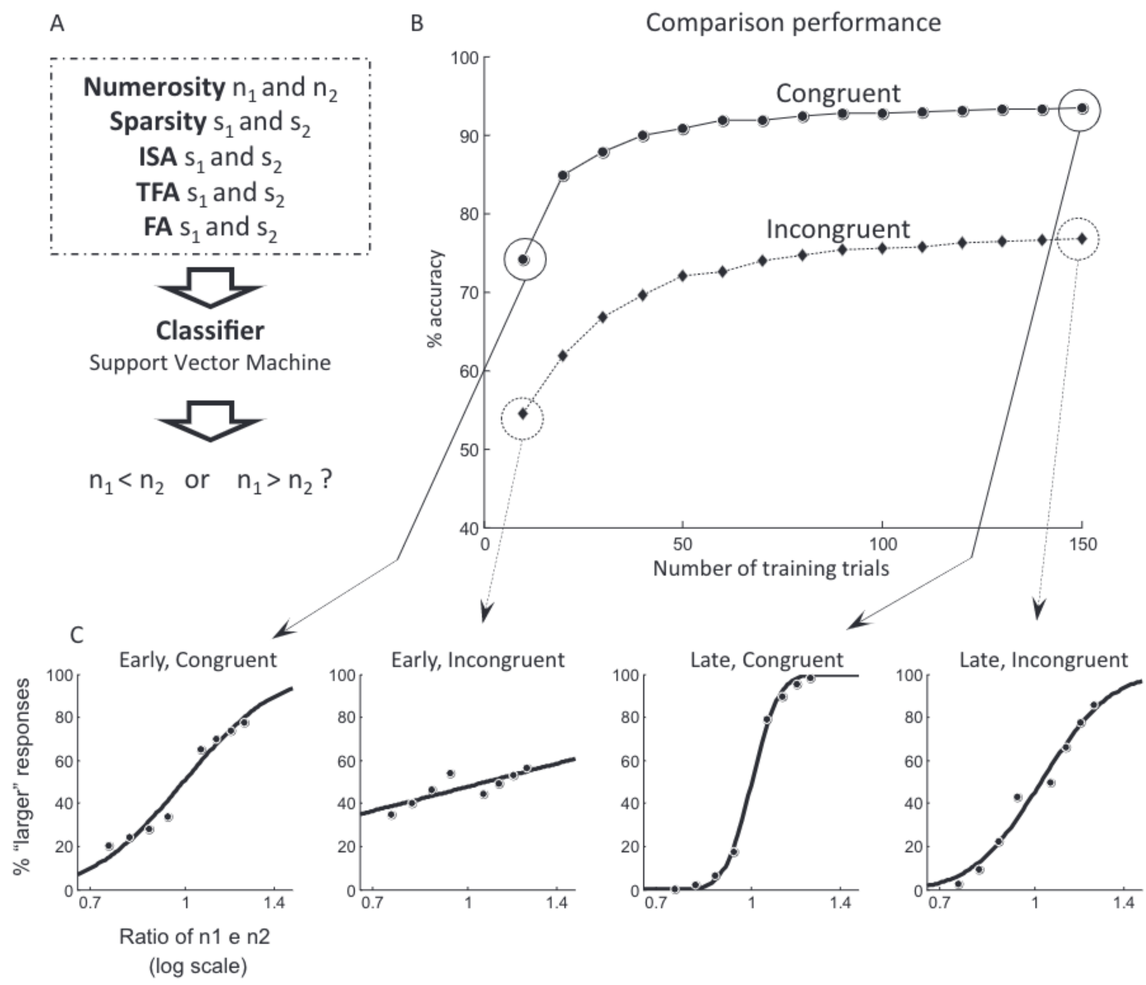


Fig. 8. Simulations of the developmental changes in numerosity comparison. A multivariate classifier (A) was trained on the problem of identifying, from the same input dimensions as in our main experiment, which of two sets contains the larger number of objects. (B) Overall performance of the classifier on congruent and incongruent trials as a function of number of training trials. (C) Performance as a function of the distance between numbers (same format as Figs. 2B and 3A), approXimating that observed in young western preschoolers and dyscalculics, as well as uneducated Mundurucú Indians vs. western educated adults.

number, but rather on its extraction by prefrontal cortex neurons and use for decision-making.

Our results fit with this animal model, but provide a step forward in several ways. First, Nieder did not contrast number to other non-nu- merical quantitative features such as length or density, but only to color information, which does not produce behavioral interference in nu- merosity tasks in humans. The present study, by contrast, addresses directly the question of the development of congruency effect between number and other quantitative non-numerical dimensions. Second, we contrasted sharpening to filtering using behavioral data and, most im- portantly, in humans subjects of variable age and education. Thus, our findings demonstrate that filtering is the principal mechanism under- lying learning not only in macaque monkeys but also in humans, and emphasize an effect of potential relevance to the field of education.

How does filtering evolve before children undertake formal math education? While simple experiments suggest that infants can sponta- neously discard density, area, and size, and respond to number (Cordes & Brannon, 2008, 2009), the precise psychophysics of numerical pre- ference and its interaction with other quantitative features during the first years of life remains to be characterized. It is possible that number and other physical variables are immediately separable by infants, but that their correlation in the natural environment is internalized through experience. This predicts that the interference effects should rapidly increase during the first months/years of life, and then decrease only later, coinciding with the onset of formal education to numerical symbols and mathematics. However, it is equally possible that number and the continuous dimensions are originally “integral” (inseparable) (Garner & Felfoldy, 1970), and become separable through experience and education (Foard & Kemler, 1984; Goldstone, 1994).

Several options are also available for the nature of the filtering process. Some authors (Cappelletti, Didino, Stoianov, & Zorzi, 2014; Houdé et al., 2011) proposed that interference effects in numerosity comparison is controlled by a central domain-general inhibitory system. This hypothesis, however, is partially questioned by the observation that the developmental trajectories of number and density judgements, which are both quantitative ensemble properties of sets that suffer from mutual interference, are different and not correlated across subjects (Odic, 2018). Another possibility is that filtering occurs without the need of effortful attentional inhibition, but that it reflects an automatic process akin to statistical learning, which continuously adjusts the weights assigned to the different dimensions to comply with the task at hand. Previous work using dual-task settings in adults showed that numerical estimation precision was not impaired by increasing the at- tentional load of a concurrent task, indirectly supporting the idea that attention is not needed to implement such filtering operation (Burr, Turi, & Anobile, 2010; Piazza, Fumarola, Chinello, & Melcher, 2011).

Our conception of this filtering mechanism also differs from a pre- vious proposal (Gilmore et al., 2013) in that we do not see it as a multi- purpose domain-general system, but rather one that specifically oper- ates in the domain of quantity judgements. Let us consider, by analogy, phonological skills and their development in relation to learning reading: with age and education, children become increasingly able to filter out word-level global phonological information and focus their attention on the syllabic sub-unit information, thus improving their "phonological awareness". Such improvements are the result of forms of filtering and inhibition that are domain specific in that they occur during reading/spelling acquisition, a very specific form of cultural training. In the same vein, we propose that another form of cultural training, that of learning to count and to operate with numerical sym- bols, acts as a powerful training tool that allows the brain to separate the signals coming from different magnitude systems, thus making numerical judgements less prone to interference from continuous magnitudes.

In conclusion, the present study provides evidence that the ability to filter out irrelevant information and amplify the relevant one is a key ingredient during perceptual learning in the number domain. Our re- sults also lead to a testable prediction for early math education: a pedagogy that systematically emphasizes the invariance of number in the face of variations and transformations in other irrelevant dimen- sions (spacing, size, density, etc.) should be the most effective.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.cognition.2018.07.011.

References

Agrillo, C., Dadda, M., Serena, G., & Bisazza, A. (2008). Do fish count? Spontaneous discrimination of quantity in female mosquitofish. Animal Cognition, 11(3), 495–503.

Bugden, S., & Ansari, D. (2015). Probing the nature of deficits in the 'ApproXimate Number System' in children with persistent developmental dyscalculia. Developmental Science, 19(5), 817–833.

Burr, D., Turi, M., & Anobile, G. (2010). Subitizing but not estimation of numerosity requires attentional resources. Journal of Vision, 10(6).

Cappelletti, M., Didino, D., Stoianov, I., & Zorzi, M. (2014). Number skills are maintained in healthy ageing. Cognitive Psychology, 69, 25–45.

Cordes, S., & Brannon, E. M. (2008). The difficulties of representing continuous extent in infancy: Using number is just easier. Child Development, 79(2), 476–489.

Cordes, S., & Brannon, E. M. (2009). The relative salience of discrete and continuous quantity in young infants. Developmental Science, 12(3), 453–463.

Cordes, S., & Brannon, E. M. (2011). Attending to one of many: When infants are sur- prisingly poor at discriminating an item's size. Frontiers in Psychology, 2, 65.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273–297.

Dehaene, S. (2007). Symbols and quantities in parietal cortex: elements of a mathematical theory of number representation and manipulation. In Attention and performance (Vol XXII, pp. 527–574). Press HU.

Dehaene, S., Izard, V., Piazza, M. (2005). Control over non-numerical parameters in numerosity experiments. Unpublished manuscript (available on http://www.unicog. org).

DeWind, N. K., Adams, G. K., Platt, M. L., & Brannon, E. M. (2015). Modeling the approXimate number system to quantify the contribution of visual stimulus features. Cognition, 142, 247–265.
Dosher, B. A., & Lu, Z.-L. (1998). Perceptual learning reflects external noise filtering and internal noise reduction through channel reweighting. Proceedings of the National Academy of Sciences, 95(23), 13988–13993.
Foard, C. F., & Kemler, D. G. (1984). Holistic and analytic modes of processing: The multiple determinants of perceptual analysis. Journal of Experimental Psychology: General, 113(1), 94–111.
Garner, W. R., & Felfoldy, G. L. (1970). Integrality of stimulus dimensions in various types of information processing. Cognitive Psychology, 1(3), 225–241.
Gebuis, T., & Reynvoet, B. (2012). The interplay between nonsymbolic number and its continuous visual properties. Journal of Experimental Psychology: General, 141(4), 642–648.
Gilbert, C. D., Sigman, M., & Crist, R. E. (2001). The neural basis of perceptual learning. Neuron, 31(5), 681–697.
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(6), e67374.
Goldstone, R. L. (1994). Influences of categorization on perceptual discrimination. Journal of Experimental Psychology: General, 123(2), 178–200.
Goldstone, R. L. (1998). Perceptual learning. Annual Review of Psychology, 49(1), 585–612.
Guillaume, M., Nys, J., Mussolin, C., & Content, A. (2013). Differences in the acuity of the ApproXimate Number System in adults: The effect of mathematical ability. Acta Psychologica (Amst), 144(3), 506–512.
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.
Harvey, B. M., Fracasso, A., Petridou, N., & Dumoulin, S. O. (2015). Topographic representations of object size and relationships with numerosity reveal generalized quantity processing in human parietal cortex. Proceedings of the National Academy of Sciences of the United States of America, 112(44), 13525–13530.

Houdé, O., Pineau, A., LerouX, G., Poirel, N., Perchey, G., Lanoë, C., ... DelcroiX, N. (2011). Functional magnetic resonance imaging study of Piaget's conservation-of- number task in preschool and school-age children: A neo-Piagetian approach. Journal of Experimental Child Psychology, 110(3), 332–346.
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.
Jordan, K., Brannon, E., Logothetis, N., & Ghazanfar, A. (2005). Monkeys match the number of voices they hear to the number of faces they see. Current Biology, 15(11), 1034–1038.
Kersey, A. J., & Cantlon, J. F. (2017). Neural tuning to numerosity relates to perceptual tuning in 3–6-year-old children. Journal of Neuroscience, 37(3), 512–522.
King, J., & Dehaene, S. (2014). Characterizing the dynamics of mental representations: The temporal generalization method. Trends in Cognitive Sciences, 18(4), 203–210.
Mante, V., Sussillo, D., Shenoy, K. V., & Newsome, W. T. (2013). Context-dependent computation by recurrent dynamics in prefrontal cortex. Nature, 503, 78–84.
Nys, J., Ventura, P., Fernandes, T., Querido, L., Leybaert, J., & Content, A. (2013). Does math education modify the approXimate number system? A comparison of schooled and unschooled adults. Trends in Neuroscience and Education, 2(1), 13–22.
Odic, D. (2018). Children's intuitive sense of number develops independently of their perception of area, density, length, and time. Developmental Science, e12533.
Panzeri, S., Senatore, R., Montemurro, M. A., & Petersen, R. S. (2007). Correcting for the sampling bias problem in spike train information measures. Journal of Neurophysiology, 98(3), 1064–1072.
Petrov, A. A., Dosher, B. A., & Lu, Z.-L. (2005). The dynamics of perceptual learning: An incremental reweighting model. Psychological Review, 112(4), 715–743.
Piazza, M., & Eger, E. (2016). Neural foundations and functional specificity of number representations. Neuropsychologia, 83.
Piazza, M., Facoetti, A., Trussardi, A. N., Berteletti, I., Conte, S., Lucangeli, D., ... Zorzi, M. (2010). Developmental trajectory of number acuity reveals a severe impairment in developmental dyscalculia. Cognition, 116(1), 33–41.
Piazza, M., Fumarola, A., Chinello, A., & Melcher, D. (2011). Subitizing reflects visuo-

spatial object individuation capacity. Cognition, 121(1), 147–153.
Piazza, M., Izard, V., Pinel, P., Le Bihan, D., & Dehaene, S. (2004). Tuning curves for approXimate numerosity in the human intraparietal sulcus. Neuron, 44(3), 547–555.
Piazza, M., Pica, P., Izard, V., Spelke, E. S., & Dehaene, S. (2013). Education enhances the acuity of the nonverbal approXimate number system. Psychological Science, 24(6), 1037–1043.
Piazza, M., Pinel, P., Le Bihan, D., & Dehaene, S. (2007). A magnitude code common to numerosities and number symbols in human intraparietal cortex. Neuron, 53(2), 293–305.
Pinel, P., Dehaene, S., Riviere, D., & LeBihan, D. (2001). Modulation of parietal activation by semantic distance in a number comparison task. Neuroimage, 14(5), 1013–1026.
Quian Quiroga, R., & Panzeri, S. (2009). EXtracting information from neuronal populations: Information theory and decoding approaches. Nature Reviews Neuroscience, 10, 173–185.
Rugani, R., Regolin, L., & Vallortigara, G. (2011). Summation of large numerousness by newborn chicks. Frontiers in Psychology, 2, 179.
Shannon, C. E. (1948). A mathematical theory of communication. Bell System Technical Journal, 27(3), 379–423.
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.
Szucs, D., Nobes, A., Devine, A., Gabriel, F. C., & Gebuis, T. (2013). Visual stimulus parameters seriously compromise the measurement of approXimate number system acuity and comparative effects between adults and children. Frontiers in Psychology, 4, 444.
Treves, A., & Panzeri, S. (1995). The upward bias in measures of information derived from limited data samples. Neural Computation, 7(2), 399–407.
Tudusciuc, O., & Nieder, A. (2009). Contributions of primate prefrontal and posterior parietal cortices to length and numerosity representation. Journal of Neurophysiology, 2984–2994.
Verguts, T., & Fias, W. (2004). Representation of number in animals and humans: A neural model. Journal of Cognitive Neuroscience, 16(9), 1493–1504.
Viswanathan, P., & Nieder, A. (2015). Differential impact of behavioral relevance on

quantity coding in primate frontal and parietal neurons. Current Biology, 25(10), 1259–1269.
Yang, T., & Maunsell, J. H. (2004). The effect of perceptual learning on neuronal responses in monkey visual area V4. The Journal of Neuroscience, 24(7), 1617–1626.