Cognitive Psychology 69 (2014) 25–45

Contents lists available at ScienceDirect

Cognitive Psychology

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

Number skills are maintained in healthy ageing

Marinella Cappelletti [a,b,*], Daniele Didino [c,1], Ivilin Stoianov [d,e,1], Marco Zorzi [d,f]

[a] UCL Institute of Cognitive Neuroscience, 17 Queen Square, London WC1N 3AR, UK
[b] Goldsmiths College, University of London, New Cross London, SE14 6NW, UK
[c] Department of Cognitive and Education Science, University of Trento, via Tartarotti 7, Rovereto, Trento, Italy
[d] Department of General Psychology, University of Padova, via Venezia 8, 35131 Padova, Italy
[e] Institute of Cognitive Sciences and Technologies, National Research Council, Rome, Italy
[f] IRCCS San Camillo Hospital, Venice-Lido, Italy

ARTICLE INFO

Article history:
Accepted 29 November 2013

Keywords:
Numerical cognition
Ageing
Numerosity perception
Number acuity
Computational modelling

ABSTRACT

Numerical skills have been extensively studied in terms of their development and pathological decline, but whether they change in healthy ageing is not well known. Longer exposure to numbers and quantity-related problems may progressively refine numerical skills, similar to what happens to other cognitive abilities like verbal memory. Alternatively, number skills may be sensitive to ageing, reflecting either a decline of number processing itself or of more auxiliary cognitive abilities that are involved in number tasks. To distinguish between these possibilities we tested 30 older and 30 younger participants on an established numerosity discrimination task requiring to judge which of two sets of items is more numerous, and on arithmetical tasks. Older participants were remarkably accurate in performing arithmetical tasks although their numerosity discrimination (also known as 'number acuity') was impaired. Further analyses indicate that this impairment was limited to numerosity trials requiring inhibiting information incongruent to numerosity (e.g., fewer but larger items), and that this also correlated with poor inhibitory processes measured by standard tests. Therefore, rather than a numerical impairment, poor numerosity discrimination is likely to reflect elderly's impoverished inhibitory processes. This conclusion is supported by simulations with a recent neuro-computational model of numerosity perception, where only the specific degradation of inhibitory processes produced a pattern that closely resembled older participants' performance. Numeracy seems therefore resilient to ageing but it is influenced by the decline of inhibitory processes

* Corresponding author. Fax: +44 (0)207 916 8517.
E-mail address: m.cappelletti@ucl.ac.uk (M. Cappelletti).
[1] These authors contributed equally to this work.

0010-0285/$ - see front matter
http://dx.doi.org/10.1016/j.cogpsych.2013.11.004

supporting number performance, consistent with the 'Inhibitory Deficit' Theory.

1. Introduction

Does our ability to use numbers and arithmetical concepts change with ageing? Are these changes specific to numeracy or do they rather reflect decline of more general cognitive processes such as attention or inhibitory processes? Numerical skills have been extensively studied in children and young adults, both in terms of development or impairment following brain lesions (Ansari, 2008; Cappelletti, 2011). However, little is known about the impact of healthy ageing on numerical skills, and the few studies that investigated this issue focused mostly on arithmetical abilities, i.e. those required when solving problems such as 8×9 or $243 + 39$. These studies concurred to show that although older participants can learn new ways to solve arithmetical problems, they show a smaller repertoire of strategies and are less efficient than younger participants in selecting among them (e.g. Duverne & Lemaire, 2005; Lemaire & Arnaud, 2008; Geary & Lin, 1998; Salthouse & Kersten, 1993), or that they do not equally engage the same brain regions as younger participants when performing arithmetical tasks (El Yagoubi, Lemaire, & Besson, 2005). However, these tasks are typically multi-componential, requiring several processes such as the retrieval of arithmetic facts, the use of procedures and the ability to monitor the steps of the problem (Cappelletti & Cipolotti, 2011). It may therefore be difficult to isolate which specific component may be affected by ageing.

An alternative approach to test the impact of ageing on numeracy skills is to assess other simpler skills (sometimes referred to as 'biologically primary skills', Geary & Lin, 1998) which are thought to be foundational to more complex, education- and language-based numerical and arithmetical abilities. One such foundational skill is thought to be our capacity to represent approximate number, which is based on encoding numerosities as analog magnitudes (e.g. Izard, Dehaene-Lambertz, & Dehaene, 2008; Stoianov & Zorzi, 2012), and relies on an 'approximate number system' (ANS, Feigenson, Dehaene, & Spelke, 2004). The ANS is often measured in terms of the ability to discriminate numerosities (e.g. which set has more elements), also referred to as 'number acuity' (Halberda, Mazzocco, & Feigenson, 2008). Number acuity is expressed as Weber fraction (*wf*), which reflects the amount of noise in the underlying approximate number representation (Halberda et al., 2008; Piazza, Izard, Pinel, Le Bihan, & Dehaene, 2004). The *wf* is highly variable across individuals (Halberda, Lya, Wilmerb, Naimana, & Germine, 2012; Halberda et al., 2008; Piazza et al., 2004), and it refines progressively from infancy to adulthood (Halberda et al., 2008; Halberda & Feigenson, 2008; Lipton & Spelke, 2003, 2012; Piazza et al., 2010). Whether it continues to improve with age is an open question: longer exposure to numbers may refine the approximate number system further, similarly to what happens to other cognitive abilities like vocabulary and semantic memory (e.g. Hedden & Gabrieli, 2004). Notably, number acuity has been found to correlate with math achievement in children (Halberda et al., 2008; Mazzocco, Feigenson, & Halberda, 2011a) and to be impaired in children with developmental dyscalculia (Mazzocco, Feigenson, & Halberda, 2011b; Piazza et al., 2010),

A few previous studies have focused on how healthy ageing participants are able to represent approximate large numerosities (i.e., more than 10 elements), which has sometimes been reported to be well maintained (e.g. Gandini, Lemaire, & Dufau, 2008; Gandini, Lemaire, & Michel, 2009; Lemaire & Lecacheur, 2007; Trick, Enns, & Brodeur, 1996; Watson, Maylor, & Bruce, 2005; Watson, Maylor, & Manson, 2002). Some of these previous results, however, are difficult to interpret. This is because in some cases the focus was mainly on the strategies used to perform the numerosity task, without reporting finer quantitative details of older participants' performance (e.g. Gandini et al., 2008). In other studies, the long presentation of the stimuli (6 s or even unlimited) may have encouraged processes different from numerosity estimation, like counting (Gandini et al., 2009; Lemaire & Lecacheur, 2007; Watson, Maylor, & Bruce, 2007). Likewise, some experimental designs did not control for continuous variables which inevitably vary when manipulating the numerosity of the display, like the total area covered by the dots, i.e. cumulative area (for discussion see Piazza et al., 2004). If these continuous variables are not taken into account, for example if in all trials an increase in numerosity always corresponds to an increase in cumulative area (e.g., Gandini et al., 2008), it is unclear whether participants judged changes in numerosity or in these continuous variables. Finally, participants' performance cannot be fully characterised when only measured as percentage of correct answers rather than in terms of finer psychophysical measures like the *wf* (Gandini et al., 2008, 2009; Lemaire & Lecacheur, 2007; Watson et al., 2005).

A recent internet-based mega-study showed a different patter of results, indicating that the ability to discriminate numerosities (as indexed by the *wf*) may indeed be sensitive to ageing (Halberda et al., 2012). This age-related deterioration, which was nevertheless not explained by the authors, may reflect either the decline of the approximate number system itself or of more peripheral cognitive processes that are involved in discriminating numerosities. For instance, working memory, attention and inhibitory processes are all critical when discriminating numerosity, for example when retrieving the quantity of a standard set of items to be compared to a test set or when inhibiting task irrelevant information, such as the area or the density of elements correlating with numerosity. Since working memory, attention and inhibitory processes tend to decline with age (Grady, 2012; Hedden & Gabrieli, 2004; Nyberg, Lövdén, Riklund, Lindenberger, & Bäckman, 2012; Salthouse, Atkinson, & Berish, 2003), they may in turn affect performance in numerosity discrimination tasks. In particular, inhibitory processes have been suggested to decline with age and to underlie age-related impairments in many cognitive functions according to the 'Inhibitory Deficit' Theory (Hasher, Lustig, & Zacks, 2007; Hasher & Zacks, 1988; Hasher, Zacks, & May, 1999). Among the separate functions of inhibition, the 'restraint function' (also called 'inhibition of dominant responses', Miyake et al., 2000) is the ability to control strong responses so that others more appropriate for the task goal can be used (Hasher et al., 2007). This function, which is sensitive to age decline (Kane & Engle, 2003; May & Hasher, 1998), may be particularly relevant in the numerosity discrimination task. This could be the case of trials where some continuous variables change orthogonally to numerosity, for instance when fewer large-size dots have a bigger cumulative area than smaller-size but more numerous dots. In this case, the task-irrelevant but salient information about cumulative area (Hurewitz, Gelman, & Schnitzer, 2006) has to be controlled in order to correctly discriminate numerosity.

Here we studied young and ageing participants in order to explore the precision of their ANS measured in terms of number acuity in a numerosity discrimination task. Our study is novel not only because it systematically measured number acuity in older people but also because it aimed to determine what underlies the pattern of spared or impaired numerosity processing. We reasoned that if number acuity does not differ between older and young participants, this may be suggestive of maintained ANS. In contrast, age-related differences in number acuity may reflect impairments specific to the number system, or alternatively decline of more general cognitive processes. To distinguish between these two possibilities, our plan was twofold: first, using established neuropsychological measures, we planned to investigate arithmetical abilities, which are thought to be linked to the ANS (Halberda et al., 2008) and are therefore expected to be impaired if the ANS is impaired. Second, using dedicated and well established tasks, we set to investigate the integrity of older participants' inhibitory processes and in particular of restraint functions which might contribute to any age-related difference in numerical abilities. Besides studying numerical abilities in older participants in depth using behavioural measures based on psychophysics and neuropsychology, a second novel aspect of our research is to combine these measures with a computational approach. This aimed to establish the type of condition that may resemble our participants' performance in the numerosity discrimination task. Our goal was to better understand whether any impairment in numerosity discrimination may be due to a global deterioration of the number system or to impaired peripheral processes, specifically those on which inhibitory functions rely on.

2. Behavioral study

2.1. Methods

2.1.1. Participants

Sixty right-handed neurologically healthy, education-matched participants with normal or corrected-to-normal vision gave written consent and were paid to participate in our study which was approved by the local research Ethics Committee. Participants were selected from the UCL Institute of Cognitive Neuroscience database because of their age: the 30 young participants had a mean age of 24.8 years (range 19–36; 13 males); the 30 older participants had a mean age of 65.77 years (range 60–75; 12 males). Ageing participants were considered neurologically normal on the basis of self-report and on their performance in the Mini Mental State Examination (MMSE, Folstein, Folstein, & McHugh, 1975; see Table 1). Information on participants' education and mathematical education in particular was collected in order to assess the possible impact of these factors on performance in the experimental tasks. Data in all participants were collected in one or two testing sessions (in the latter case, they were 7–10 days apart).

2.2. Numerosity discrimination: experimental tasks and stimuli

Stimulus presentation and data collection were controlled using the Cogent Graphics toolbox (http://www.vislab.ucl.ac.uk/Cogent) and MATLAB 7.3 software on a Sony S2VP laptop computer with video mode of 640 $\times$ 480 pixels, and 60 Hz refresh rate.

Following earlier studies that investigated number acuity across the lifespan (Halberda et al., 2012), we used a version of the numerosity discrimination task that required to judge which of two intermixed collections of coloured dots was more numerous. Sets of dots were presented in blue and yellow in each display, with 5–16 dots for each colour. The ratios between the larger and the smaller number of dots were 2:1, 4:3, 6:5 and 8:7, with 40 trials presented for each of the easiest ratios (i.e. 2:1 and 4:3), and 120 for each of the most difficult ratios (6:5 and 8:7) (see Table S1). We used a larger number of trials in the most difficult ratios in order to increase statistical power (i.e., to obtain more reliable estimates of individual *wf*) since we expected lower accuracy in these ratios. A total of 320 trials was presented in 10 blocks, and in each ratio the 'larger' set was equally assigned to the two colours.

For each ratio there was also an equal number of congruent (dot-size controlled) and incongruent (area controlled) trials presented in random order. *Dot-size controlled* trials were those in which the average diameter of the dots in the larger set was equal to the average diameter of the smaller set. The diameter of a dot ranged approximately between 0.57° and 1.17° of visual angle from a distance of 57 cm, the average diameter being 0.87°. In these trials the cumulative area of the larger set was always larger than the cumulative area of the smaller one. In contrast, *area controlled* trials were those in which the average diameter of the larger set (which ranged approximately between 0.57° and 1.17° of visual angle, i.e. 0.87 ± 35%) was smaller than the average diameter of the smaller set (which ranged approximately between ±35% of the average diameter of the smaller set itself). The average diameter of the smaller set was selected so that the cumulative area of the two sets was equal.

2.2.1. Procedure

Each trial started with a fixation point for 1500 ms followed by a display of blue and yellow dots for 200 ms after which a question mark appeared to prompt participants to respond (see Fig. 1). Participants were instructed to make an unspeeded answer indicating which group of dots (blue or yellow) was more numerous by pressing one of two predefined computer keys (M or N), whose assignment was randomised between participants. Once an answer was made, the following trial started immediately with the 1500 ms fixation.

2.2.2. Data analysis

For all data, the Shapiro–Wilk test confirmed the normality of the distribution, and the data were analysed using parametric tests (multiple regressions, Analyses of Variance (ANOVAs) and *t*-tests). The data sphericity was tested using the Mauchy Test, and for significant results Greenhouse–Geisser correction was applied in case of sphericity violation. Significance was set at a p value of 0.05.

Following earlier studies, we fitted the response distribution of each participant in the numerosity discrimination task to obtain individual estimates of number acuity (i.e., the precision of the underlying numerical representation), expressed as internal Weber fraction (*wf*, e.g. Halberda et al., 2012, 2008; Piazza et al., 2004, 2010. See SI for a detailed description of how the *wf* was calculated). Although speed was not stressed in the instructions given to participants, we analysed response times

Table 1
Older and young participants' (A) demographic information, and performance in (B) background tasks, (C) number and arithmetic tasks, (D) executive and inhibitory processes. Percentile, reaction time (RTs), percent correct or Weber Fraction (*wf*) with standard deviation (SD).

Task/information	Young participants ($N = 30$)	Older participants ($N = 30$)
A. Demographic information		
Age	24.8 years (range 19–36)	65.77 years (range 60–75)
Gender	13 males	12 males
Years of education	18	16
Years of mathematical education	13	12
B. Background		
Full IQ (WAIS-R)[a]	116.2 (10.8)	118.3 (13.4)
Mini Mental State Examination[b]	nt	29.4 (0.6)
Stimulus identification	293.1 ms (33.72)	348.2 ms (34.0)
Vocabulary[a]	83%ile (18.5)	94%ile (6.3)
Digit span[a]	82%ile (19.6)	86%ile (12.3)
Door recognition[c]	87%ile (15.1)	83%ile (14.7)
C. Number and arithmetic		
Numerosity discrimination based on		
All trials (*wf*)	0.24 (0.04)	0.30 (0.07)
Congruent trials (*wf*)	0.238 (0.06)	0.255 (0.08)
Incongruent trials (*wf*)	0.256 (0.06)	0.367 (0.15)
Number comparison		
Accuracy (% correct)	97.1 (2.5)	98.9 (1.6)
RTs small/large distance[d]	576 ms (141)/488 ms (76)	698 ms (112)/607 ms (65)
Arithmetic tasks		
Arithmetic verification		
Accuracy (% correct); RTs	94.2 (4.4); 1127 ms (241)	97.3 (2.9); 1409 ms (387)
Graded Difficulty Arithmetic Test[e]	80.6%ile (18.2)	86.3%ile (19.9)
WAIS-R math sub-test[a]	83.1%ile (16.4)	86.3%ile (22.8)
Reading and writing numbers	99.5% (1.9)	98.7% (2.1)
D. Executive and inhibitory processes		
Number Stroop task		
Magnitude comparison		
Congruent	99.3%; 478 ms (67)	99.8%; 745 ms (99)
Neutral	97.17%; 514 ms (64)	99.6%; 789 ms (112)
Incongruent	89.1%; 544 ms (66)	96.4%; 866 ms (120)
Physical comparison		
Congruent	99.1%; 457 ms (71)	99.8%; 600 ms (76)
Neutral	99.8%; 470 ms (62)	99.8%; 615 ms (79)
Incongruent	93.1%; 505 ms (82)	99%; 687 ms (84)
Word Stroop task[f]		
Word target		
Congruent	98.7%; 462 ms (55)	99.3%; 602 ms (82)
Neutral	99.7%; 472 ms (48)	99.3%; 612 ms (72)
Incongruent	93.3%; 513 ms (68)	98.0%; 696 ms (72)
Colour target		
Congruent	98.3%; 467 ms (102)	99.3%; 611 ms (103)
Neutral	95.6%; 473 ms (114)	99.8%; 610 ms (93)
Incongruent	94.8%; 470 ms (80)	98.8%; 725 ms (161)
Attention network test (ANT)[g]		
Congruent; Incongruent	448 ms (86); 533 ms (59)	658 ms (82); 766 ms (90)
Conflict: incongruent–congruent	85 ms (20)	98 ms (26)

nt = not tested; ms = milliseconds.
[a] Wechsler (1995).
[b] Folstein et al. (1975); max score: 30.
[c] Baddeley A. D., H., & I., 1994.
[d] Longer RTs to small number distances indicate normal distance effect [$r = 0.4$, $F(1,29) = 14.8$, $p < 0.001$], not different from young participants [$t(58) = 8.4$, $p < 0.001$].
[e] Jackson and Warrington (1986).
[f] Stroop, 1935.
[g] Fan et al. (2002).

to further characterise participants' performance following other studies using a similar experimental paradigm and procedure (e.g. Halberda et al., 2012; Piazza et al., 2010).

2.2.3. Results of numerosity discrimination

Accuracy differed across the four numerosity ratios [percent correct, ratio 2:1, Young = 96% (sd = 0.4), Older = 89% (sd = 0.7); ratio 4:3, Young = 78% (sd = 0.9), Older = 77% (sd = 0.6); ratio 6:5, Young = 71% (sd = 0.4), Older = 66% (sd = 0.5); ratio 8:7, Young = 65% (sd = 0.6), Older = 62% (sd = 0.4)].

An analysis of the *wf* in young participants indicated a large variability, such that the *wf* was on average 0.24, ranging from 0.16 to 0.33, consistent with previous studies (e.g. Halberda et al., 2008, 2012). Likewise, number acuity in older participants showed large individual differences, being on average 0.30 and ranging from 0.20 to 0.49.

In a regression analysis with *wf* as dependent variable and group as independent one, we found that group was a significant predictor of number acuity [beta = 0.38, SE = 0.16, $t = 3.1$, $p = 0.003$], as confirmed by a significant difference between older and young participants [$t(58) = 3.2$, $p = 0.002$, see Fig. 2 left panel]. A more specific analysis looking at the variability of performance in the older sample showed that within the older group 21 out of 30 participants showed a *wf* that was one or two standard deviations below the average of the younger group, indicating defective performance. This suggests that the group difference we observed was not simply driven by a few participants, thereby mirroring the results of Halberda et al.'s (2012) internet mega-study.

We also tested whether education in general and mathematical education in particular may predict number acuity. This is because mathematical expertise (e.g. in accountants and bookkeepers) has previously been shown to explain differences in memory ability for arbitrary numbers and for grocery-store prices between expert and non-expert ageing participants (Castel, 2005, 2007). Although our

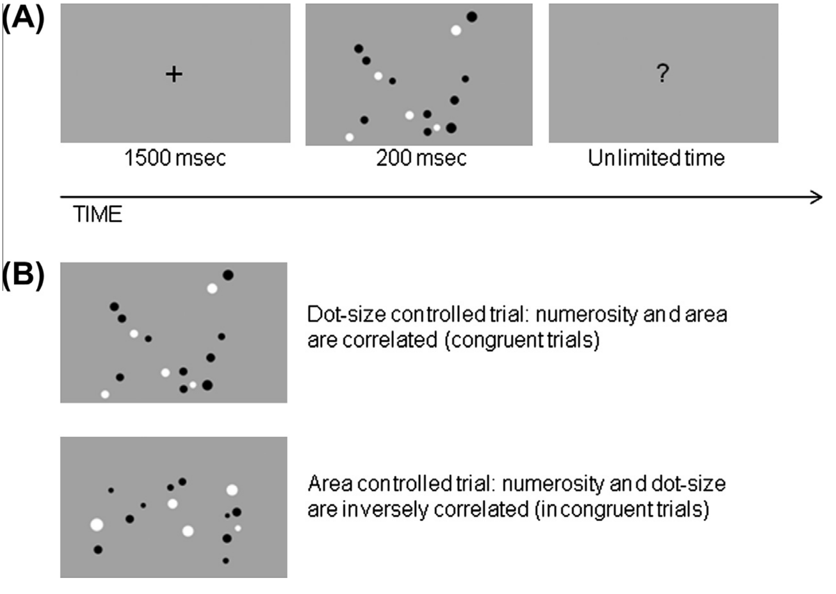

Fig. 1. Numerosity discrimination task. (A) Participants saw a centrally presented fixation cross for 1500 ms immediately followed by a brief (200 ms) presentation of a display of yellow and blue dots (for the purpose of the figure displayed in black and white). Participants were instructed to indicate with a button press whether each trial contained more blue or yellow dots, with no time constrains. The following trial was displayed immediately after the answer. (B) Example of area-controlled (incongruent) and dot-size controlled (congruent) trials. Participants were presented with 320 trials; half of these trials contained yellow and blue dots having on average the same cumulative area and whose diameter was inversely correlated to numerosity, implying that the least numerous set had the larger dot diameter (incongruent trials, bottom panel of B). The other half of the trials contained yellow and blue dots of approximately the same average diameter such that the larger display had larger cumulative area (congruent trials, top panel of B).

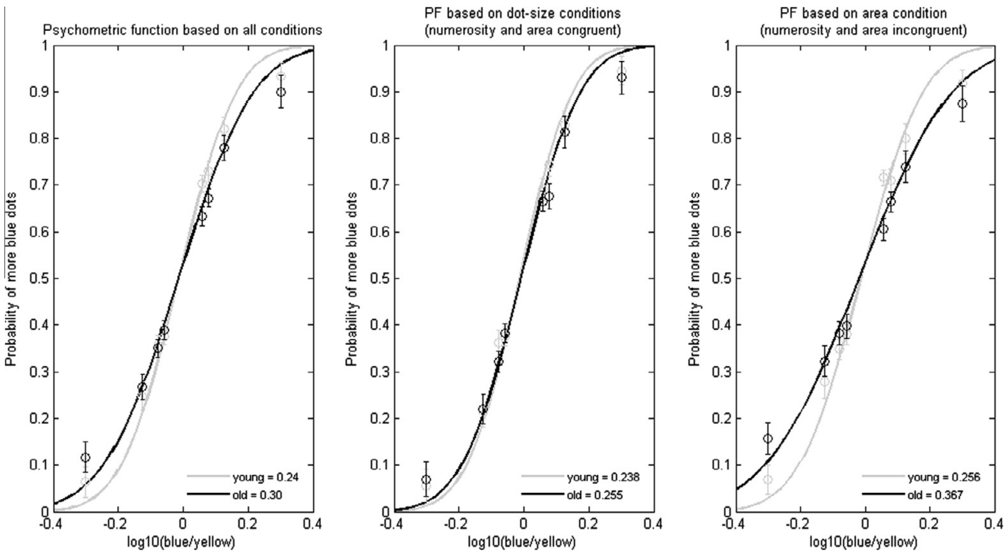

Fig. 2. Behavioural results. Psychometric functions (PF) indicating young (in light gray) and older (in black) participants' performance in the numerosity discrimination task measured in terms of *wf* in: (A) all conditions (i.e. average of dot-size and area-controlled trials), (B) dot-size controlled trials only, and (C) area-controlled trials only. The psychometric functions plot the probability of choosing 'more blue' dots (*y*-axis) as a function of the ratio between blue and yellow dots (*x*-axis).

sample did not include mathematical experts, we nevertheless explored whether education and mathematical education may in part explain the difference in number acuity we observed. However, we found that neither of these predictors were significant [education: beta = 0.08, SE = 0.004, $t = 0.58$, $p = 0.6$; mathematical education: beta = 0.009, SE = 0.004, $t = 0.06$, $p = 0.9$].

The difference in *wf* might suggest that relative to younger participants, older participants needed a larger discrepancy between the two stimulus sets to be able to identify the larger, which may be due to a deteriorated representation of numerosity.

2.3. Arithmetical tasks

Impaired performance in the numerosity discrimination task may reflect a deteriorated representation of numerosity, which has been suggested to be linked to arithmetical skills, following the proposal that number acuity is foundational to these skills (Halberda et al., 2008). We therefore tested older and younger participants' arithmetical abilities, predicting a possible impairment in older participants if these skills are linked to ANS.

2.3.1. Stimuli and methods for the arithmetical tasks

Stimuli for the arithmetical tasks were Arabic numbers in the form of single to 3 digits presented either on the computer, on paper or read aloud by the experimenter. We used the following four tasks: arithmetic verification, number comparison, multi-digit mental arithmetic, and arithmetical problem solving.

Arithmetic verification required participants to indicate as fast as possible whether an arithmetic problem displayed the correct or incorrect answer (using the 'F' or 'J' keys of the keyboard). Twenty single-digit problems for each type of operation (addition, subtraction and multiplication) were presented in separate blocks with no repetition of the same items in immediately successive trials. All problems had operands below 10; results of the addition and subtraction problems were between 1 and 18, and of multiplication problems were between 6 and 36. Following a 500 ms central fixation cross, each operation was presented for up to 7 s during which participants could answer. When presented with incorrect results, these were either 1 or 2 units apart from the correct result for addition and subtraction problems (e.g. 9 + 6 = 13 or 7–2 = 3) or 2 units apart for multiplication problems (e.g. $6 \times 3 = 16$).

Number Comparison required indicating as fast as possible the larger of two Arabic numbers presented to the left and right of a central fixation point (using the 'F' and 'J' keys of the keyboard to indicate the larger number appearing on the left or on the right respectively). Thirty-six pairs of single-digit Arabic numbers (1–9) were individually presented.[2] Stimuli pair were centred along the horizontal line of the computer screen and displayed for 500 ms each to the left or the right of the fixation cross; they were replaced by a black screen for a maximum of 2500 ms during which participants made an answer. After this, the next trial started immediately. The following numerical distances were used: 1 (e.g. 7 vs 8), 2 (e.g. 3 vs 1), 3 (e.g. 5 vs 2), 4 (e.g. 1 vs 5), 5 (e.g. 4 vs 9). The larger number in each pair was equally presented to the left and to the right; likewise, correct answers were equally assigned to the left or the right digit in each pair.

Multi-digit mental arithmetic was tested with the Graded Difficulty Arithmetical Test (GDA, Jackson & Warrington, 1986), a standardised test based on two separate blocks of twelve 2 to 3-digit addition and twelve 2 to 3-digit subtraction problems of progressive difficulty (e.g. from '13 + 15' to '243 + 149'). These were orally presented one at a time for an oral answer, which scored 1 point if correctly produced within 10 s.

Arithmetical problem solving was assessed with the arithmetical subtest of the Wechsler Intelligence Scale (WAIS-R; Wechsler, 1995), consisting of twenty arithmetical problems embedded in a text and orally presented for an oral answer. Correct answers produced within a maximum time (spanning from 15 to 60 s depending on the problem) were assigned 1 point.

Performance in these numerical and arithmetical tasks was expressed as accuracy and/or reaction times (RTs) of correct answers only; reaction times were collected in the number comparison and the arithmetic verification tasks only.

Since performance in arithmetical skills are often more sensitive to speed than accuracy, and since speed tends to increase with age (e.g. Salthouse, 1991), we measured older participants' general speed of performance in two ways. The first way was based on a simple perceptual task, which required participants to respond as rapidly as possible to a dot stimulus appearing on the computer monitor at various locations. Forty trials were presented, each displaying a dot for 200 ms in random locations on the left or right of a computer monitor and following an ISI randomly selected between 500 ms and 2 s. Participants were instructed to make speeded responses whenever the stimulus appeared by pressing a pre-defined key. Accuracy and response times were recorded.

The second way consisted of measuring older participants' speed of performance using a two-choice non-numerical task whose modality of response resembled the type of choice participants had to make in the number and arithmetical tasks. Specifically, we used responses made in the Stroop 'neutral' conditions of the Word Stroop task (averaged across the case where the target colour was presented on the neutral string 'XXX', and where the target word was presented in the neural colour grey; see also Section 2.4.2.4).

2.3.2. Results of arithmetical tests

Remarkably, we found that young and older participants accuracy did not significantly differ in any of the number and arithmetical tasks (all $P > 0.4$, see Table 1). In older participants, proficiency in the numerosity task correlated with accuracy in the arithmetical tasks [arithmetic verification: $r = 0.51$, $F(1,29) = 9.7$, $p < 0.004$; multi-digit arithmetic (GDA): $r = 0.42$, $F(1,29) = 6.1$, $p < 0.02$; arithmetic problem solving (WAIS): $r = 0.36$, $F(1,29) = 4.2$, $p < 0.05$], but not in number comparison [$r = 0.15$, $F(1,29) = 0.6$, $p = 0.4$, *ns*].

Despite maintained accuracy, older participants were significantly slower than young to perform the number comparison and the arithmetic verification tasks [$t(58) = 13.1$, $p < 0.001$ and $t(58) = 3.2$,

[2] The pairs of Arabic numbers were as follows; distance 1: 1–2, 3–2, 4–3, 5–4, 6–7, 7–8, 8–9; distance 2: 3–1, 2–4, 4–6, 5–3, 7–5, 6–8, 9–7; distance 3: 4–1, 2–5, 3–6, 7–4, 8–5, 6–9, 5–2, 1–4; distance 4: 1–5, 3–7, 4–8, 5–9, 6–2, 7–3, 8–4, 9–5; distance 5: 1–6, 2–7, 8–3, 9–4.

$p < 0.01$ respectively). We therefore tested whether this slowness was specific for the number tasks or whether instead it reflected general slowness. We run a fixed-entry hierarchical regression with the basic RTs, the choice RTs and group entered as independent variables in different steps, and RTs (separate for arithmetic verification and number comparison) as the dependent variable. Results showed a significant effect of group [arithmetical verification: beta = −275.6, SE = 133.5, $t = 2.0$, $p < 0.05$; number comparison: beta = −245.6, SE = 49.2, $t = 4.9$, $p < 0.001$], suggesting that older participants were generally slower than younger. However, a non-significant effect of RTs in the basic perceptual speed task [arithmetical verification: beta = 1.8, SE = 1.0, $t = 1.7$, $p = 0.09$, *ns*; number comparison: beta = 0.5, SE = 0.38, $t = 1.3$, $p = 0.18$, *ns*] suggests that basic RTs did not explain participants' slowness in the number and arithmetic tasks. Instead, a significant effect of two-choice RTs [arithmetic verification: beta = 1.9, SE = 0.39, $t = 4.8$, $p < 0.001$; number comparison: beta = 0.32, SE = 0.14, $t = 2.2$, $p = 0.03$] which contributed to 65% and 88% of the variance in the two tasks respectively, accounted for participants' slowness in the number and arithmetic tasks.

2.4. The role of other cognitive processes in number acuity

Our results suggest that despite a larger *wf*, older participants' numerical and arithmetical skills were well maintained. This therefore does not support one of our initial hypotheses that a larger *wf* may reflect a more general impairment in number and arithmetical processing.

An alternative possibility is that a larger *wf* may be explained by impairments in inhibiting task-irrelevant information, an hypothesis that has been previously put forward to account for elderly's impaired performance in other cognitive functions (Hasher & Zacks, 1988; Hasher et al., 1999, 2007). If this is the case, differences between congruent and incongruent numerosity trials may be expected, since incongruent trials required inhibiting task-irrelevant information coming from continuous variables such as the size of the dots that negatively correlated with numerosity (Dakin, Tibber, Greenwood, Kingdom, & Morgan, 2011; Gebuis & Reynvoet, 2012; Hurewitz et al., 2006). In addition, we also examined older participants' inhibitory skills independently from number acuity.

2.4.1. Congruent vs incongruent numerosity trials

To investigate whether older participants' larger *wf* may reflect a difficulty to inhibit information that may be irrelevant to numerosity judgments, separate *wf* for congruent (dot-size controlled) and incongruent (area controlled) trials in the numerosity discrimination task were calculated and compared within and between groups using paired and independent *t*-tests respectively. Relative to congruent trials, incongruent numerosity trials resulted in a larger *wf* in both groups (older: congruent = 0.255 and incongruent = 0.367; young: congruent = 0.238 and incongruent = 0.256), but with a significant difference between trial types only in the older [$t(29) = 3.9$, $p < 0.001$] but not in the young participants [$t(29) = 0.9$, $p = 0.3$, *ns*], see Fig. 2 middle panel. The *wf* differed between the two groups only in the incongruent [$t(58) = 3.7$, $p < 0.001$] but not in the congruent trials [$t(58) = 0.1$, $p = 0.4$, *ns*; see Fig. 2 right panel].

In older participants, age correlated with performance in the incongruent numerosity trials [$r = 0.3$, $F(1,29) = 2.9$, $p < 0.04$] but not with the congruent ones [$r = 0.12$, $F(1,29) = 0.4$, $p = 0.5$, *ns*].

2.4.2. Numerosity discrimination task with blocked congruent and incongruent trials

Having found that older participants were specifically impaired in the incongruent numerosity trials as indexed by a significantly larger *wf* relative to young participants, we aimed to identify which factors may account for such impairment. One possibility is that older participants' impaired incongruent numerosity trials might be related to the strategies used to perform these trials. For instance, in order to identify the more numerous set older participants may have used a strategy based on just numerosity, and also other strategies, for example based on a salient non-numerical visual feature (like the area covered by the dots or amount of colour as a proxy for numerosity). These numerosity and area-based strategies may have been successfully used in the case of incongruent and congruent trials respectively, but switching between them may have been particularly difficult for ageing participants. We reasoned that a difficulty in switching among strategies may be ameliorated by keeping the congruent and incongruent numerosity trials blocked. We therefore administered participants the same numerosity task but with the congruent and incongruent trials in blocked rather than randomized order, and with a trial-by-trial feedback to help maintaining correct performance. Since we only found a large *wf* in the older participants, we did not test younger participants in this blocked and feedback-based version of the numerosity task.[3]

2.4.2.1. Stimuli and method. We used the same stimuli, design and procedure previously employed (see Section 2.2), with the following changes. First, the same number ($N = 320$) and type of trials previously used were now presented in random order in 5 blocks of 32 congruent trials and 5 blocks of 32 incongruent trials. The 5 blocks of congruent and the 5 blocks of incongruent trials were grouped and the order was counterbalanced across participants. Second, to encourage correct performance, trial-by-trial feedback was introduced at the end of each trial, indicating whether participants made a correct or wrong answer. All the other features of the task were exactly the same as before (see Section 2.2). Performance was calculated in terms of *wf* with the same procedure previously used; we also measured response times, although speed was not stressed in the instructions to participants.

2.4.2.2. Results. The *wf* and response times obtained in the numerosity task with blocked and with the random trial presentation were compared in independent *t*-tests. There was no significant advantage in presenting trials in a blocked fashion and with feedback relative to the canonical random presentation, indicated by an almost unchanged *wf* (average random condition: 0.29 vs blocked condition: 0.31; $t(11) = 0.3$, $p = 0.7$, *ns*). This suggests that it is unlikely that higher *wf* in older participants is simply due to problems in switching among strategies, for instance from a numerosity-based to an area-based rule. Further support to this suggestion comes from the analysis of the response times, which were examined in an ANOVA with presentation type (blocked vs random) and trial type (congruent vs incongruent) as within-subject factor. This showed no main effect of presentation [$F(1,11) = 0.8$, $p = 0.3$, *ns*] and of trial type [$F(1,11) = 0.7$, $p = 0.4$, *ns*], and no significant interaction [$F(1,11) = 0.02$, $p = 0.8$, *ns*], therefore suggesting no difference between the blocked and the random conditions.

2.4.2.3. Analysis of response times in the numerosity task. The pattern of response times in the numerosity discrimination task (data from the main experiment, see Section 2.2) was analysed to assess the hypothesis that group differences may be due to a speed-accuracy trade-off for the older participants. We used a multivariate analysis of variance (MANOVA) with response times and accuracy (*wf)* for the congruent and incongruent trials entered as dependent variables, and group (young and older) and type of trial (congruent and incongruent) as fixed factors. Both group and trial type were significant [respectively: $F(2,115) = 16.04$, $p < 0.001$, Wilk's $\Lambda = 0.78$; and $F(2,115) = 15.05$, $p < 0.001$, Wilk's $\Lambda = 0.79$], as well as their interaction [$F(2,115) = 4.8$, $p < 0.01$, Wilk's $\Lambda = 0.9$]. Specifically, group had a significant impact on both accuracy [$F(1,116) = 7.3$, $p < 0.001$] and response times [$F(1,116) = 26.3$, $p < 0.001$], and likewise trial type significantly affected accuracy [$F(1,116) = 21.9$, $p < 0.00$] and response times [$F(1,116) = 9.7$, $p = 0.002$]. Group and trial type also interacted with accuracy [$F(1,116) = 5.1$, $p < 0.02$] and with response times [$F(1,116) = 4.1$, $p = 0.04$]. This is because young participants were more accurate than older in incongruent (*wf*: 0.256 vs 0.367; $t(58) = 3.7$, $p < 0.001$) but not in congruent trials (*wf*: 0.238 vs 0.255; $t(58) = 0.1$, $p = 0.4$, *ns*), and response times for congruent and incongruent numerosity trials differed in the older [617 ms vs 649 ms, $t(29) = 2.3$, $p = 0.02$] but not in the young participants [415 ms vs 427 ms, $t(29) = 0.9$, $p = 0.3$, *ns*]. This RTs significant difference in trial type in the ageing group was maintained even when older participants' general slowness was taken into account in a regression analysis. Here response times for the congruent and incongruent numerosity trials were separately entered as dependent variable in the analysis, with group and two-choice RTs as independent variables. There was a significant effect of group in the incongruent trials only [congruent: beta = 0.3, SE = 65.7, $t = 1.5$, $p = 0.13$, *ns*; incongruent: beta = 3.9, SE = 71.8, $t = 2.0$, $p = 0.048$], suggesting that older participants were slower than younger only in the incongruent trials. In these trials there was no main effect of choice RTs [beta = 0.25, SE = 0.22, $t = 1.2$, $p = 0.2$, *ns*], indicating that slowness in performing incongruent trials could not be accounted for by general

[3] Only 12 older participants could be tested on this additional version of the numerosity task.

slowness, although this seemed to be case for congruent trials where choice RTs were marginally significant [beta = 3.5, SE = 0.2, $t = 1.8$, $p = 0.06$]. These findings suggest that older participants did not trade off their accuracy for response speed. Indeed, both accuracy and RT data revealed a specific difficulty of older participants in processing incongruent trials.

These results suggest that larger *wf* in older participants were due to differences in *wf* in the incongruent but not the congruent trials, possibly because these trials relied upon inhibitory skills. We therefore aimed to explore these skills in more details in our older participants. We focused on inhibition not only because it is critical for processing the incongruent numerosity trials but also because it has been suggested to decline with age and to underlie age-related impairments in many other cognitive functions (Hasher & Zacks, 1988; Hasher et al., 1999, 2007).

2.4.2.4. Older participants' performance in inhibitory tasks. Using standard tests, we specifically focused on the restraint function among inhibitory processes, i.e. the ability to control strong responses such that others more appropriate for the task goal can be used (Hasher, Lustig, & Zacks, 2007). We reasoned that this function may be particularly relevant in the numerosity discrimination task and specifically in the incongruent trials where information about the area covered by the dots had to be controlled in order to correctly discriminate numerosity. We used the following standard tasks to assess inhibitory processes in older participants: Word Stroop, Number Stroop, Attention Network test (ANT).

The word Stroop task requires participants to read as quickly as possible either a word ignoring the colour of the ink it is printed on (for instance 'RED' whether printed in colour red, blue or in grey for a neutral condition), or to name the colour in which words are printed ignoring their meaning (for instance to name the colour red whether displayed on the word 'RED', 'BLUE' or on 'XXX' for a neutral condition). There were 60 trials for each task (word or colour). For the word task, stimuli were the word 'RED' or 'BLUE', which could appear in red, blue or gray colour; this therefore resulted in a congruent, incongruent or neutral condition (20 trials for each type), depending on whether, for example, the word 'RED' appeared in colour red, or in colour blue or in gray. In the colour task, stimuli were the word 'RED', 'BLUE' or a string of 'XXX'. These could appear in colour red or blue, therefore resulting in a congruent, incongruent or neutral condition (20 trials for each type), depending on whether, for example, the colour red appeared on the word RED or on the word BLUE or on the string XXX.

In each trial, participants saw a centrally presented 500 ms fixation cross, followed by a word stimulus displayed until the participant made an answer or for a maximum of 4000 ms. After this, the following trial started immediately. Participants were asked to decide as quickly as possible whether the stimulus was the word 'RED' or 'BLUE' irrespective of the colour (in the word task), or whether it was displayed in red or blue font irrespective of the meaning of the word (colour task); they were instructed to press the left and right arrow keys for blue (word or colour) and red (word or colour) respectively. The two tasks were presented separately, and the order of the tasks was counterbalanced across participants.

For each task (word or colour), accuracy and response times were calculated for the three conditions: neutral (corresponding to the target word printed in grey or the target colour printed on XXX), congruent (target word printed on the corresponding colour or the target colour on the corresponding word), and incongruent (target word printed on a different colour like 'RED' printed in blue, or the target colour printed on a different word, like colour red on the word blue). These three conditions allowed calculating an index of 'congruity' (i.e. response times of correct answers only in incongruent–congruent trials), a standard measure of participants' ability to inhibit task irrelevant information (Stroop, 1935).

The number Stroop task is based on an established paradigm that assesses the automatic processing of numbers as well as inhibitory processes using experimental stimuli that contain congruent and incongruent information (Henik & Tzelgov, 1982). In two separate tasks, participants viewed a total of 336 pairs of 1–9 Arabic numbers (168 per block) that could vary in magnitude (e.g. 3 vs 2) or physical size (e.g. 3 vs 2). There were therefore three types of stimuli (each of 36 trials per task): a congruent stimulus corresponded to a pair of digits in which a given digit was larger in both the relevant and the irrelevant dimensions; a neutral stimulus was a pair of digits that differed only on the relevant dimension (magnitude or physical size); an incongruent stimulus consisted of a pair of digits in which one of the digits was at the same time larger in one dimension (e.g. magnitude) and smaller in the other (e.g. physical size). Each number stimulus could be paired to itself, therefore consisting of a neutral stimulus for the physical size condition (e.g. 2 vs 2), or to another number stimulus which could be between 1 and 4 units apart. Moreover, the two number stimuli could be of the same physical size, therefore consisting of the neutral stimulus for the magnitude condition, or they could vary along two levels of physical size, as stimuli could appear in a vertical visual angle of 0.7° or 0.9° centred along the horizontal line of the computer screen to the left or the right of the fixation cross.

Participants were required to indicate the larger number in either magnitude or physical size by pressing either the left or the right arrow if the larger number was presented either to the left or to the right. Following a 500 ms fixation cross, the number stimuli were presented until the participant made an answer or for a maximum of 4000 ms. After this, the next fixation cross appeared and the following trial started immediately. For each task (magnitude or physical size), accuracy and response times were recorded. Using this experimental design, it is common to find a congruity effect which consists of facilitation, i.e. faster responding to stimuli in which information about the magnitude and the physical size of the stimuli is congruent (e.g. 3 vs 2) relative to neutral trials (e.g. 3 vs 3 for physical comparisons or 3 vs 2 for numerical comparisons), and of interference namely slower responses to stimuli in which information about the magnitude and the physical size of the stimuli is incongruent (e.g. 3 vs 2) relative to neutral trials (Henik & Tzelgov, 1982). These effects are thought to reflect automatic processing of numerical and physical size even when magnitude or size are task-irrelevant; this is suggested by studies showing that irrespective of which dimension is relevant, response times for congruent trials are always shorter than response times for incongruent trials, and for incongruent trials shorter than for neutral trials, although an advantage for congruent relative to neutral trials has only been observed in the magnitude condition (Henik & Tzelgov, 1982; Tzelgov, Meyer, & Henik, 1992). This task therefore served the double purpose of measuring participants' automatic number processing and, critically, also their ability to inhibit task-irrelevant information.

The Attention Network Test (ANT, Fan, McCandliss, Sommer, Raz, & Posner, 2002) examines executive and inhibitory processes using a variation of the flanker task (Eriksen & Eriksen, 1974) whereby participants attend to one object while ignoring others (Posner, 1980). In the version used here, a cuing task and a flanker task were combined such that participants responded to cued or un-cued central targets while ignoring flanking distractors. A total of 288 trials were presented in 3 blocks of 96 trials each. The stimuli consisted of a target arrow flanked by two arrows on either side, which could appear in the same direction as the target arrow (congruent condition e.g. ← ← ← ← ←) or in the opposite direction (incongruent condition e.g. ← ← → ←←). Following Fan and colleagues' paradigm (2002), each arrow was presented at 0.55° of visual angle and separated from the adjacent arrows by 0.06° of visual angle. The stimuli (the central arrow and the four flankers) consisted of a total 3.08° of visual angle. Participants were instructed to attend to the middle arrow and to decide whether it was pointing to the left or to the right. A trial consisted of the following events: a central fixation cross was first presented for a random duration between 400 and 1600 ms, and followed by either a 100 ms warning asterisk cue in the cued trials or by a longer fixation in the un-cued trials, and by a second 400 ms fixation period after which the target and the flankers appeared simultaneously and centrally, at 1.06° of visual angle either above or below the fixation point. The cue could appear centrally, hence corresponding to a spatially neutral condition or it could precede the target and flankers in the same position (the cue was always valid), i.e. at 1.06° of visual angle above or below the fixation point, which corresponded to a spatially orienting condition. The target and flankers remained on the screen until the participant responded or for a maximum of 1700 ms. After a response was made, the next trial began immediately.

Participants had to press as quickly as possible a left-hand key if the central arrow pointed left and a right-hand key if it pointed right. The task allows measuring three indexes of performance based on how response times of correct answers are influenced by alerting cues, spatial cues, and flankers: alertness (cued vs un-cued trials), orienting (central cue vs spatial cue), and conflict (congruent vs incongruent trials averaged across cued and un-cued, and central vs spatial cue). As conflict is the critical component we aimed to measure, we limited the analysis of the performance to this index of behaviour.

2.4.2.5. Results of tasks assessing inhibitory processes. Older participants significantly differed from younger participants in the ability to suppress task-irrelevant information (see Table 1). In both tasks of the Word Stroop (i.e. where the target was either a word or a colour), older participants were significantly slower in incongruent than congruent trials relative to young (word Stroop task: $t(58) = 3.2$, $p = 0.002$; colour Stroop task: $t(58) = 7.5$, $p < 0.001$). Likewise in both the Number Stroop tasks and in the ANT, older participants showed a larger congruity effect relative to young participants [Number Stroop tasks: magnitude comparison, $t(58) = 3.7$, $p < 0.001$; physical size comparison, $t(58) = 2.1$, $p < 0.04$; ANT, $t(58) = 2.7$, $p = 0.008$]. Accuracy was near ceiling in both groups (98.6% and 98.9% correct in older and young respectively), and did not differ between groups [$t(58) = 0.76$, $p = 0.54$, *ns*].

Critically, older participants' poor performance in these tasks correlated with *wf* in the incongruent numerosity trials. Hence, the worse the elderly's performance in these incongruent relative to congruent trials, the longer were their response times in the incongruent trials of the Stroop tasks and of the Attention Network Test [Word Stroop (averaged over word and colour target conditions): $r = 0.38$, $p < 0.04$; Number Stroop (averaged over magnitude and physical conditions): $r = 0.4$, $p = 0.03$; ANT: $r = 0.37$, $p = 0.04$]. This was not the case of performance in the congruent trials which did not correlate with any measure of inhibition processes [Word Stroop: $r = 0.24$, $p < 0.2$, *ns*; Number Stroop: $r = 1.5$, $p = 0.4$, *ns*; ANT: $r = 0.06$, $p = 0.7$, *ns*]. The equivalent analyses in young participants showed no link between *wf* in congruent or incongruent numerosity trials and response times in any of the above tasks [all $P > 0.1$].

These results suggest a link between *wf* in the incongruent trials and older participants' performance in tests of inhibitory processes. This in turn may reflect deteriorated inhibitory processes rather than impoverished number processing, as this would have likely resulted in impaired performance in other arithmetical tasks. To further support the hypothesis that larger *wf* may reflect deteriorated inhibitory processes, we performed a set of computer simulations based on a recent neurocomputational model of numerosity perception (Stoianov & Zorzi, 2012).

3. Computational modelling study

The connectionist model of numerosity perception developed by Stoianov and Zorzi (2012) is a deep neural network with two hierarchically organised layers of (hidden) neurons. The first layer receives input from the image and it consists of uniformly spread center-On local detectors (high frequency spatial filters). The second layer consists of local numerosity detectors that receive input from the center-On neurons normalised by an inhibitory signal representing the image cumulative area. The activity of numerosity detectors (layer 2 neurons) is used as input to a simple linear classifier that is trained to decide whether the input numerosity is greater than a reference number. Performance of the model in the numerosity comparison task is evaluated in the same way as human participants, that is by computing the model's number acuity (*wf*) from the accuracy data.

The main aim of our simulations was to establish what type of impairment would produce the dissociation between intact performance on dot-size controlled ("*congruent*") trials and declined performance on area-controlled ("*incongruent*") trials observed in the elderly population. Rather than fitting the participants' performance, the simulation aimed to offer an explanation at the level of the mechanisms that subtend numerosity discrimination in the model. To provide a closer match to the ANS task used in the present behavioral study, the model was extended to simulate the comparison of two variable visual numerosities.

3.1. Computer simulations

3.1.1. Stimuli

Using the method described in Stoianov and Zorzi (2012), we created visual numerosity images of size 30×30 pixels, each containing up to 32 randomly placed non-overlapping rectangular objects of variable size and shape[4] (see samples in Fig. 3). Object size varied within each pattern according to a Gaussian term with mean = 0 and sd = 0.5. We created a database of images containing objects with

[4] Note that the range of numerosities (with upper limit of 32 objects) was selected by Stoianov and Zorzi (2012) to allow the simulation of the typical experimental setting of numerosity comparison studies (e.g., Piazza et al., 2004). This requires a minimum image size of 30×30 pixels to keep all objects spatially separated by at least one pixel in the largest numerosity condition.

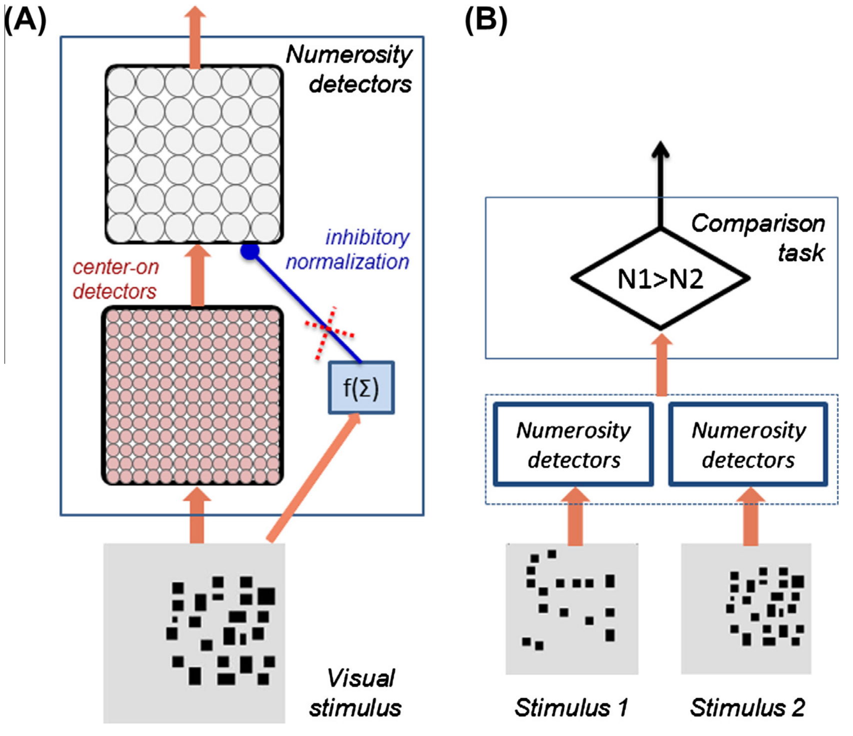

Fig. 3. Computational model. (A) Hierarchical network for numerosity perception (Stoianov & Zorzi, 2012). The input image (30 × 30 pixels) is processed by one layer of center-on detectors with small-size receptive fields. Numerosity detectors at the top layer have larger receptive fields and compute a local numerosity signal by combining the activity of center-on detectors with an inhibitory normalization signal conveying the image's cumulative area (see text for details). Poor numerosity discrimination in the elderly population can be accounted for by degraded inhibitory normalization. (B) Numerosity comparison task. Activity of the numerosity detectors obtained for each image of a pair was used as input to a linear classifier, that is a simple network trained to decide which of the two numerosites was larger.

variable cumulative area that ranged from 32 to 256 pixels at 8 levels, with 200 images for each level of numerosity ($n = 32$) and cumulative area ($n = 8$). We then compiled a training set by drawing 10240 pairs of images from the image database, ensuring that numerosity and cumulative area were balanced. The training set was used to train the linear classifier on the numerosity comparison task, while the model's performance (i.e., *wf*) was assessed using an independent set of test images that were not used for training (as in Stoianov & Zorzi, 2012).

We also created a database of test images that had the same characteristics of the stimuli used in Experiment 1. In particular, area-controlled test images had constant cumulative area of 100 pixels, whereas size-controlled test images contained objects with constant size (3 × 3 pixels). Cumulative area was not predictive of numerosity in the area-controlled images but it was correlated with the numerosity in the size-controlled ones. The database contained 200 images for each of the numerosities and conditions (size-controlled vs. area-controlled) of the behavioural study. The test set was created by drawing 1100 pairs of test images for each of the two conditions, yielding the same numerosity pairs (and ratios) used in the behavioural study.

3.1.2. Model details

The simulations extended the simplified version of the model described in the Supplementary Material of Stoianov and Zorzi (2012) where the numerosity perception network is hard-wired and training involves only the numerosity discrimination task. The model is depicted in Fig. 3. It is worth noting that the architecture and the parameters of the model (e.g., size of the receptive fields) were identical to those reported in the original study and they were chosen as to faithfully represent the structure emerged (i.e., self-organised) in the learning model. For the present purposes, a key advantage of the simplified mathematical model is that it clearly identifies the functionality of each model component, such as the role of inhibitory signals (see below).

The numerosity perception network has two hierarchically organised layers of neurons above the visual input layer. The first layer (size: 13 × 13 neurons) receives input from a binary image I and it consists of uniformly spread center-On local detectors (i.e. high frequency spatial filters) activated according to the following equation:

$$O^{ij} = f\left(\sum V^{ij} I + b\right) \tag{1}$$

where V^{ij} are 2D-gaussian-shaped receptive fields ($\sigma = 2$) over the image I, b is a bias term (fixed to −1), and $f(.)$ is the logistic function. The second layer (size: 6 × 6 neurons) consists of uniformly spread local numerosity detectors activated according to the following equation:

$$N^{kl} = \sum W^{kl} O - c \tag{2}$$

where W^{kl} are 2D-Gaussian-shaped receptive fields ($\sigma = 6$) over the center-ON neurons and c is a (inhibitory) normalisation term based on the image cumulative area,

$$c = \log\left(1 + \frac{\sum I}{c_{max}}\right) \tag{3}$$

where c_{max} is the maximal cumulative area across the image database. The activity of the numerosity detectors coarsely represents numerosity. Indeed, Stoianov and Zorzi (2012) showed that the response profile of the population activity of layer 2 numerosity neurons closely resembled the monotonic coding of numerosity observed in monkey LIP neurons (Roitman, Brannon, & Platt, 2007) and that it was invariant to perceptual properties of the stimuli, such as cumulative area.

Stoianov and Zorzi (2012) showed that the activity of numerosity detectors supported human-like performance in numerosity comparison when used as input to a linear classifier that was trained to compare numerosities to a fixed reference. More specifically, the activity of the numerosity detectors elicited by a given input image was fed to a decision layer encoding a "smaller" vs. "larger" judgment. This simple linear network was trained using the delta rule, which is formally equivalent to the Rescorla-Wagner learning rule and has been widely used to account for human learning (Sutton & Barto, 1981). Thus, the model assumes that numerosity comparison is a simple linear task when the input to the decision is an abstract representation of numerosity. This is line with Dehaene and Changeux (1993) computational model of the development of early numerical abilities and it fits well with the observation that the classic numerosity comparison paradigm (with explicit judgment/response) can be carried out even by 3–4 year old children (Halberda et al., 2008; Piazza et al., 2010). The model responses, when plotted as a function of numerical ratio, yielded a psychometric function that perfectly mirrored those of adult observers, thereby obeying Weber's law. Moreover, the *wf* computed from the model's response distribution was virtually identical to the mean value reported for adult observers (e.g., Piazza et al., 2010). The model's descriptive adequacy was therefore supported by its match to both behavioral (i.e., human psychophysics) and neural (i.e., single cell recording) data (see Stoianov & Zorzi, 2012, for further details).

Here we simulated the comparison of two variable numerosities (as in the behavioral paradigm of Experiment 1 and in the studies of Halberda et al., 2008, 2012) by feeding two internal numerosity representations (i.e., the activity of the numerosity detectors for each image of a given pair in the training set) to a linear classifier that was trained to decide which of the two numerosities was larger. Learning was based on the delta rule and the training database was presented for 100 epochs (when learning became markedly asymptotic). Model's performance after training (i.e., model's *wf*) was assessed on the image pairs contained in the test set. Because the representation of each numerosity (activity of numerosity detectors) is intrinsically noisy, we expected greater response variability (i.e., a higher *wf*) in this version of the task relative to the original simulations of Stoianov and Zorzi (2012) where a single numerosity was compared to a fixed reference number.

3.2. Elderly model

To simulate the declined performance of elderly population, we implemented a degradation to the network connections by applying a stochastic decay term to the synaptic strengths (i.e., weight values). This stochastic reduction of synaptic efficacy results in neurons' decreased responsivity to afferent signals. Decreased responsivity of neurons, which turns into a weaker signal-to-noise ratio (i.e., increased neuronal noise) is also critical in the neuromodulation model of ageing of Li and colleagues (Li, Lindenberger, & Bäckman, 2010; Li, Lindenberger, & Sikström, 2001). In their model, this is obtained by flattening the slope of the neurons' sigmoid activation function, which is contolled by a stochastic gain parameter. Li and colleagues link the value of the gain parameter to dopaminergic modulation, which is known to be altered in elderly population.

We implemented two different types of impairment: a global degradation of the numerosity perception network, which involved all network synapses, or a selective degradation of the inhibitory synapses (in line with the inhibition deficit hypothesis). Inhibition in the model is critical for abstracting numerosity from continuous visual properties, specifically cumulative area (see Eq. (2)). Indeed, the normalisation signal represents an inhibitory input to the numerosity detectors that is conveyed through feed-forward inhibitory connections (see Fig. 3). Feed-forward inhibition is known to act as input normalisation mechanism in real neurons (e.g., Pouille, Marin-Burgin, Adesnik, Atallah, & Scanziani, 2009). In the present model, the excitatory input to numerosity detectors increases both with numerosity and cumulative area (because large objects will tend to activate more center-on detectors than small objects) but is normalised by an inhibitory input whose strength also increases with cumulative area (see Stoianov & Zorzi, 2012, for detailed analyses).

Synaptic degradation was obtained by scaling the connection weights by a random value (range [0–1]) drawn from a Gaussian distribution with mean = m and s.d. = 0.1 (smaller values of m determine stronger impairment). Numerosity comparison in the model was adjusted after the impairment through a short re-training (5 epochs) on a new dataset created just as the training set and containing 5120 image pairs. The re-training was meant to capture a gradual adjustment of the decision process to compensate for the aging effect. Indeed, aging is known to modify the decision criteria in two-choice decision tasks (Ratcliff, Thapar, & McKoon, 2006). The severity of each type of impairment, controlled by the parameter m, was set to yield an average decline in performance (across 30 replications) that matched the decline observed in human participants. This procedure yielded a global impairment characterised by $m = 0.20$ and, alternatively, an inhibitory impairment with $m = 0.25$.

We performed 30 replications of each type of impairment (i.e. global vs. inhibitory) and we assessed the model's performance in numerosity comparison on the test dataset. Individual *wfs* were separately computed for the two conditions (size-controlled vs. area-controlled) of the test dataset. A control population of 30 unimpaired networks was created and assessed after applying the same readjustment procedure used for the impaired networks (to ensure that the observed differences were not due differences in training regimen). Test and re-training dataset were independently drawn for each replication.

4. Results of the computational model

The *wf* values obtained in the simulation (computed for each replication) were submitted to a mixed ANOVA with condition (area-controlled vs. dot size-controlled) as within-subject factor and impairment type (unimpaired, inhibitory, global) as between-subject factor. Both factors, and importantly, their interaction, affected numerosity discrimination (impairment: $F_{(2,87)} = 26$, $p < 0.001$; condition: $F_{(1,87)} = 212$, $p < 0.001$; interaction: $F_{(2,87)} = 187$, $p < 0.001$). Planned contrasts revealed that both types of impairment worsened discrimination performance relative to the unimpaired models, with overall *wf* (i.e., across conditions) increasing from 0.249 to 0.313 after global impairment ($t(58) = 9.8$, $p < 0.001$) and to 0.315 after inhibition impairment ($t_{(58)} = 5.5$, $p < 0.001$), see Fig. 4. Performance of the two types of impaired model was not statistically different ($t_{(58)} = 0.2$, $p = 0.85$). Critically, the inhibition impairment specifically worsened discrimination on area-controlled stimuli ($wf = 0.384$) relative to dot-size controlled stimuli ($wf = 0.246$; $t_{(29)} = 20.6$, $p < 0.001$). Global

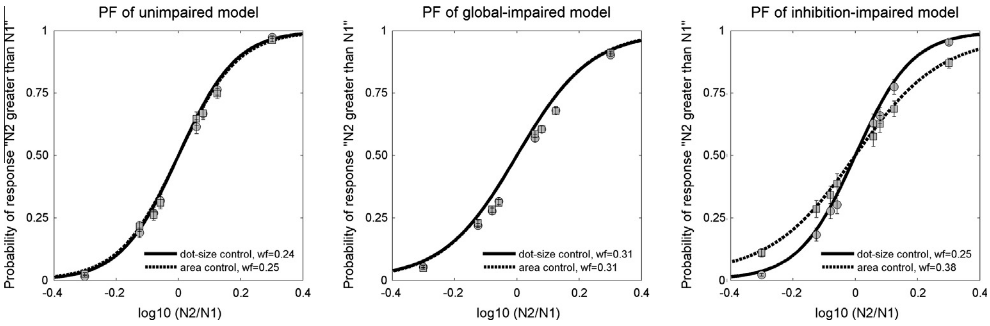

Fig. 4. Computational model results. Psychometric functions (PF) indicating the model's numerosity discrimination performance (averaged across networks) for each type of impairment (unimpaired, global and inhibition impaired) and test dataset (area-controlled vs. size-controlled).

impairment yielded similar discrimination performance in both conditions (area-controlled: *wf* = 0.311, dot-size controlled: *wf* = 0.315; $t_{(29)} = 0.5$, $p = 0.59$). The unimpaired models showed a performance comparable to that of our young adult participants in both conditions, with a slight advantage on the congruent (*wf* = 0.244) relative to the incongruent (*wf* = 0.253) stimuli ($t_{(29)} = 2.1$, $p = 0.044$). Finally, the inhibition impairment deteriorated performance (relative to that of the unimpaired models) in the incongruent condition ($t_{(29)} = 8.6$, $p < 0.001$) but not in the congruent condition ($t_{(29)} = 0.2$, $p = 0.85$).

In sum, just as in the elderly human participants, specific impairment of inhibition caused a large decrease of performance on stimuli in which task-relevant and irrelevant features compete. Conversely, an equally strong global impairment caused a decline in performance that was identical across conditions. Reduced inhibition of irrelevant information is therefore critical to explain the specific pattern of numerosity discrimination performance in elderhood.

5. Discussion

We investigated whether numerical and arithmetical skills may be impaired in healthy ageing. We first studied whether normally ageing participants had impaired number acuity – which tests a foundational numerical skill measured in terms of Weber fraction, *wf* – and found that older participants showed a significantly larger *wf* relative to young participants. This is in line with the results of Halberda et al.'s internet mega-study (Halberda et al., 2012), and may indicate impaired numerosity processing. However, a novel and more detailed analysis showed that this larger *wf*, i.e. impaired numerosity processing, could be due to a specific inability to process trials in which numerosity was incongruent with other measures of continuous quantity (i.e., where the least numerous set contained dots of larger size), whilst congruent trials (i.e., those where the more numerous set had a larger cumulative area) were performed equally accurately in both populations. This impairment in incongruent trials did not correspond to difficulties in other numerical or arithmetical problems which were exceptionally well maintained in older participants.

Whether older participants performed the congruent numerosity trials by processing just numerosity or whether they successfully combined information about numerosity with information about other continuous dimensions like the area covered by the dot stimuli may be difficult to establish. Indeed, even if numerosity can effectively be extracted in the context of other dimensions such as cumulative area or dot size (e.g., Stoianov & Zorzi, 2012), information from these dimensions can influence numerosity judgments (Gebuis & Reynvoet, 2012; Hurewitz et al., 2006; Stoianov & Zorzi, submitted for publication). However, older and younger participants' *wf* were very similar in the congruent numerosity trials therefore suggesting that either both groups consistently used an area-based strategy to perform the numerosity task, or alternatively that they both mainly used a numerosity-based strategy. The first possibility seems unlikely as participants would have had to continuously switch between strategies (e.g. between an area-based and a non area-based strategy), which seems inefficient, and was not corroborated by differences in response times. Moreover, in older participants this switching might have resulted in faster response times in the numerosity task with blocked conditions, which we did not find. The second possibility is that both groups used a numerosity-based strategy which had to be efficiently combined with inhibitory processes in the numerosity incongruent trials. We suggest that ageing participants may have used a numerosity-based strategy to perform the numerosity task and that their larger *wf* can be explained by impoverished inhibitory processes.

This hypothesis was supported by the results of tasks assessing inhibitory processes independent from numerosity. We found that elderly's performance in these tasks was significantly different from younger participants and that it correlated with performance in the incongruent numerosity trials. Therefore, the more difficult it was to inhibit task-irrelevant information, the worse was the performance in the incongruent numerosity trials. We further tested the hypothesis that performance in incongruent numerosity trials may depend on the integrity of inhibitory processes using a computational model. This showed that a general degradation of the number system resulted in a larger *wf* with no difference between dot-size and area-controlled trials. However, a more specific degradation of the inhibitory processes resulted in a significantly larger impairment in processing the area-controlled numerosity trials, which resembled our older participants' performance.

Approximating larger numerosities (i.e. more than 10 elements) or 'number acuity' (Halberda et al., 2008) has been shown to be either well maintained (Gandini et al., 2008; Trick et al., 1996; Watson et al., 2005) or conversely impaired in older participants (Halberda et al., 2012). However, different aims or sub-optimal methodological aspects of some of these previous studies make it difficult to interpret some of the reported results. Similar to these earlier studies, we aimed to investigate the pattern of spared or impaired numeracy skills in elderly participants, but we also sought to provide further and more specific information on what exactly lies behind elderly's number performance. For the first time, our results of ageing participants' maintained performance in congruent numerosity trials but impaired in incongruent ones points at two fundamental processes intrinsic to numerosity discrimination: the abstraction of numerosity in a set and the inhibition of task-irrelevant information. We suggest that the first of these two processes was maintained in the ageing group as indicated by an unchanged *wf* in numerosity congruent trials relative to younger participants and by maintained arithmetical performance. In contrast, inhibitory processes were partially defective in ageing participants and we suggest that this is what affected performance in the incongruent numerosity trials. Such distinction between abstraction and inhibitory processes in numerosity discrimination is likely to be undetected in younger populations, or even in elderly if the experimental design does not allow distinguishing between congruent and incongruent trials, or if performance measures do not allow finer analyses.

Inhibitory processes have so far been suggested to account for elderly's impaired performance in several processes, for instance memory and language, based on the 'Inhibitory Deficit' Theory (Hasher et al., 2007; Healey, Campbell, & Hasher, 2008). Here we suggest that these processes may also account for the impoverished performance we observed in the numerosity trials that required inhibitory processes in order to be accurately judged. It may also be argued that in ageing participants the impairment lies at the encoding level (Stoltzfus, Hashe, & Zacks, 1996). That is, older participants may not have encoded numerosity information properly and this in turn may account for the larger *wf*. However, this seems less likely because an encoding problem, which would make information difficult to retrieve, would have also affected the congruent numerosity trials.

Our novel, combined approach allowed us to avoid the misleading conclusion that ageing affects numerosity discrimination. Instead, intact number acuity and arithmetical skills in ageing participants adds to previous observations that despite cognitive decline in some functions, the ageing brain is able to maintain other skills like vocabulary, syntax and semantic memory (e.g. Hedden & Gabrieli, 2004). However, differently from these education-related skills whose maintenance may partly be due to increasing practice with age, undamaged number abilities are likely to reflect the preservation of a system that is at least partially unrelated to education and language and that is thought to be robustly embedded in the human brain (Feigenson et al., 2004). This is in line with the evidence that quantification processes, albeit in a more rudimental form, are already present in children and infants (Lipton & Spelke, 2003), and are relatively better maintained than other processes following brain lesions (Cappelletti, Butterworth, & Kopelman, 2012). The idea that numerical and arithmetical skills are resilient to normal ageing because they rely on a primitive and approximate number system (see Hasher & Zacks, 1979 for a similar view), is consistent with the idea that primitive systems tend to be more robust to ageing (Lemaire & Lecacheur, 2007; Trick et al., 1996). This is because primitive systems tend to be innate or to be acquired earlier in life and this may put them in a stronger position relative to more complex and recently acquired skills (Trick et al., 1996). It is also possible that the number system relies on brain areas that are naturally less affected by ageing (Nyberg et al., 2012) or where decline begins later, a working hypothesis for future studies.

Acknowledgments

This work was supported by a Royal Society Dorothy Hodgkin Fellowship, a Royal Society and a British Academy Research Grants (M.C), and by Grant No. 210922 from the European Research Council (M.Z.). We thank Brian Butterworth for his comments on an earlier version of the manuscript.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.cogpsych.2013.11.004.

References

Ansari, D. (2008). Effects of development and enculturation on number representation in the brain. *Nature Reviews Neuroscience, 9*, 278–291.
Baddeley A. D., Emslie, H., & Nimmo-Smith, I. (1994). *Doors and people: A test of visual and verbal recall and recognition*. Bury St. Edmunds, England: Thames Valley Test Company.
Cappelletti, M., & Cipolotti, L. (2011). The neuropsychological assessment and treatment of calculation disorders. In Halligan, Kischka, & Marshall (Eds.). *Handbook of clinical neuropsychology*. Oxford University Press.
Cappelletti, M., Butterworth, B., & Kopelman, M. (2012). Numerical abilities in patients with focal and progressive neurological disorders: A neuropsychological study. *Neuropsychology, 26*, 1–19.
Castel, A. D. (2005). Memory for grocery prices in younger and older adults: The role of schematic support. *Psychology and Aging, 20*, 718–721.
Castel, A. D. (2007). Aging and memory for numerical information: The role of specificity and expertise in associative memory. *Journal of Gerontology: Psychological Sciences, 62*, 194–196.
Dakin, S., Tibber, M., Greenwood, J., Kingdom, F. A., & Morgan, M. (2011). A common visual metric for approximate number and density. *Proceedings of the National Academy of Sciences of the United States of America, 108*, 19552–19557.
Dehaene, S., & Changeux, J. P. (1993). Development of elementary numerical abilities: A neuronal model. *Journal of Cognitive Neuroscience, 5*, 390–407.
Duverne, S., & Lemaire, P. (2005). Arithmetic split effects reflect strategy selection: An adult age comparative study in addition verification and comparison tasks. *Canadian Journal of Experimental Psychology, 59*, 262–278.
El Yagoubi, R., Lemaire, P., & Besson, M. (2005). Effects of aging on arithmetic problem solving: An event-related brain potential study. *Journal of Cognitive Neuroscience, 17*, 37–50.
Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identication of a target letter in a nonsearch task. *Perception & psychophysics, 16*, 143–149.
Fan, J., McCandliss, B. D., Sommer, T., Raz, A., & Posner, M. I. (2002). Testing the efficiency and independence of attentional networks. *Journal of Cognitive Neuroscience, 14*, 340–347.
Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences, 8*, 307–314.
Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12*, 189–198.
Gandini, D., Lemaire, P., & Dufau, S. (2008). Older and young adults' strategies in approximative quantification. *Acta Psychologica, 129*, 175–189.
Gandini, D., Lemaire, P., & Michel, B. F. (2009). Approximate quantification in young, healthy older adults, and Alzheimer patients. *Brain and Cognition, 70*, 53–61.
Geary, D. C., & Lin, J. (1998). Numerical cognition: Age-related differences in the speed of executing biologically primary and biologically secondary processes. *Experimental Aging Research, 24*, 101–137.
Gebuis, T., & Reynvoet, B. (2012). The role of visual information in numerosity estimation. *PLoS One, 7*, e37426.
Grady, C. (2012). The cognitive neuroscience of ageing. *Nature Review Neuroscience, 20*, 491–505.
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*, 1457–1465.
Halberda, J., Lya, R., Wilmerb, J. B., Naimana, D. Q., & Germine, L. (2012). Number sense across the lifespan as revealed by a massive Internet-based sample. *Proceedings of the National Academy of Sciences of the United States of America*. http://dx.doi.org/10.1073/pnas.1200196109.

Halberda, J., Mazzocco, M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature, 455*, 665–668.
Hasher, L., Lustig, C., & Zacks, R. T. (2007). Inhibitory mechanisms and the control of attention. In A. Conway, C. Jarrold, M. Kane, A. Miyake, & A. J. Towse (Eds.), *Variation in working memory* (pp. 227–249). New York: Oxford University Press.
Hasher, L., & Zacks, R. T. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General, 108*, 356–388.
Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. In G. H. Bower (Ed.). *The psychology of learning and motivation* (Vol. 22, pp. 193–225). New York, NY: Academic Press.
Hasher, L., Zacks, R. T., & May, C. P. (1999). Inhibitory control, circadian arousal, and age. In D. Gopher & A. Koriat (Eds.), *Attention & performance, XVII, cognitive regulation of performance: Interaction of theory and application* (pp. 653–675). Cambridge, MA: MIT Press.
Healey, M. K., Campbell, K. L., & Hasher, L. (2008). Cognitive aging and increased distractibility: Costs and potential benefits. In W. S. Sossin, J. C. Lacaille, V. F. Castellucci, & S. Belleville (Eds.). *Progress in brain research* (Vol. 169, pp. 353–363). Amsterdam: Elsevier.
Hedden, T., & Gabrieli, J. (2004). Insight into the aging mind: A view from cognitive neuroscience. *Nature Reviews Neuroscience, 5*, 87–97.
Henik, A., & Tzelgov, J. (1982). Is three greater than five: The relation between physical and semantic size in comparison tasks. *Memory & Cognition, 10*, 389–395.
Hurewitz, F., Gelman, R., & Schnitzer, B. (2006). Sometimes area counts more than number. *Proceedings of the National Academy of Sciences of the United States of America, 103*, 19599–19604.
Izard, V., Dehaene-Lambertz, G., & Dehaene, S. (2008). Distinct cerebral pathways for object identity and number in human infants. *PLoS Biology, 6*(2), e11.
Jackson, M., & Warrington, E. K. (1986). Arithmetic skills in patients with unilateral cerebral lesions. *Cortex, 22*, 611–620.
Kane, M. J., & Engle, R. W. (2003). Working memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General, 132*, 47–70.
Lemaire, P., & Arnaud, L. (2008). Young and older adults' strategies in complex arithmetic. *American Journal of Psychology, 121*, 1–16.
Lemaire, P., & Lecacheur, M. (2007). Aging and numerosity estimation. *Journal of Gerontology: Psychological Sciences, 62*, P305–P312.
Li, S.-C., Lindenberger, U., & Bäckman, L. (2010). Dopaminergic modulation of cognition across the life span. *Neuroscience & Biobehavioral Reviews, 34*, 625–638.
Li, S.-C., Lindenberger, U., & Sikström, S. (2001). Aging cognition: From neuromodulation to representation. *Trends in Cognitive Sciences, 5*, 479–486.
Lipton, J. S., & Spelke, E. S. (2003). Origins of number sense: Large number discrimination in human infants. *Psychological Science, 15*, 396–401.
May, C. P., & Hasher, L. (1998). Synchrony effects in inhibitory control over thought and action. *Journal of Experimental Psychology: Human Perception and Performance, 24*, 363–379.
Mazzocco, M. M. M., Feigenson, L., & Halberda, J. (2011a). Preschoolers' precision of the approximate number system predicts later school mathematics performance. *PLoS One, 6*(9), e23749.
Mazzocco, M. M. M., Feigenson, L., & Halberda, J. (2011b). Impaired acuity of the approximate number system underlies mathematical learning disability (dyscalculia). *Child Development, 82*, 1224–1237.
Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41*, 49–100.
Nyberg, L., Lövdén, M., Riklund, K., Lindenberger, U., & Bäckman, L. (2012). Memory aging and brain maintenance. *Trends in Cognitive Sciences, 16*, 292–305.
Piazza, M., Facoetti, A., Trussardi, A. N., Berteletti, I., Conte, S., Lucangeli, D., et al (2010). Developmental trajectory of number acuity reveals a severe impairment in developmental dyscalculia. *Cognition, 116*, 33–41.
Piazza, M., Izard, V., Pinel, P., Le Bihan, D., & Dehaene, S. (2004). Tuning curves for approximate numerosity in the human intraparietal sulcus. *Neuron, 44*, 547–555.
Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology, 41A*, 19–45.
Pouille, F., Marin-Burgin, A., Adesnik, H., Atallah, B. V., & Scanziani, M. (2009). Input normalization by global feedforward inhibition expands cortical dynamic range. *Nature Neuroscience, 12*, 1577–1585.
Ratcliff, R., Thapar, A., & McKoon, G. (2006). Aging and individual differences in rapid two-choice decisions. *Psychonomic Bulletin and Review, 13*, 626–635.
Roitman, J., Brannon, E. M., & Platt, M. L. (2007). Monotonic coding of numerosity in macaque lateral intraparietal area. *PLoS Biolology, 5*, e208.
Salthouse, T. A. (1991). Mediation of adult age differences in cognition by reductions in working memory and speed of processing. *Psychological Science, 2*, 179–183.
Salthouse, T. A., Atkinson, T. M., & Berish, D. E. (2003). Executive functioning as a potential mediator of age-related cognitive decline in normal adults. *Journal of Experimental Psychology: General, 132*, 566–594.
Salthouse, T. A., & Kersten, A. W. (1993). Decomposing adult age differences in symbol arithmetic. *Memory & Cognition, 2*, 699–710.
Stoianov I., & Zorzi, M. (submitted for publication). The influence of continuous visual properties on numerosity estimation: No challenge for the Approximate Number System.
Stoianov, I., & Zorzi, M. (2012). Emergence of a "visual number sense" in hierarchical generative models. *Nature Neuroscience, 15*, 194–196. http://dx.doi.org/10.1038/nn.2996.
Stoltzfus, E. R., Hashe, L., & Zacks, R. T. (1996). Working memory and aging: Current status of the inhibitory view. In J. T. E. Richardson, R. W. Engle, L. Hasher, R. H. Logie, E. R. Stoltzfus, & R. T. Zacks (Eds.), *Counterpoints in cognition: Working memory and human cognition* (pp. 66–68). Oxford, UK: Oxford University Press.
Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643–662.

Sutton, R. S., & Barto, A. G. (1981). Toward a modern theory of adaptive networks: Expectation and prediction. *Psychological Review, 88*, 135–170.

Trick, L. M., Enns, J. T., & Brodeur, D. A. (1996). Life span changes in visual enumeration: The number discrimination task. *Developmental Psychology, 32*, 925–932.

Tzelgov, J., Meyer, J., & Henik, A. (1992). Automatic and intentional processing of numerical information. *Journal of Experimental Psychology: Learning, Memory and Cognition, 18*, 166–179.

Watson, D. G., Maylor, E. A., & Bruce, L. A. M. (2005). Effects of age on searching for and enumerating targets that cannot be detected efficiently. *Quarterly Journal of Experimental Psychology, 58*, 1119–1142.

Watson, D. G., Maylor, E. A., & Bruce, L. A. M. (2007). The role of eye movements in subitizing and counting. *Journal of Experimental Psychology: Human Perception and Performance, 33*, 1389–1399.

Watson, D. G., Maylor, E. A., & Manson, N. J. (2002). Aging and enumeration: A selective deficit for the subitization of targets among distractors. *Psychology and Aging, 17*, 496–504.

Wechsler, D. (1995). *Wechsler abbreviated scale of intelligence*. *The Psychological Corporation*. New York, NY: Harcourt Brace & Company.