The handle http://hdl.handle.net/1887/26495 holds various files of this Leiden University dissertation.

**Author:** Davidse, Neeltje Joanne  
**Title:** Links between executive functions and early literacy and numeracy  
**Issue Date:** 2014-06-25
Finally, by means of FEA, stronger inferences about causality can be made, but estimates of the effects are more imprecise due to larger standard errors. Thus, although FEA reduces the risk of misleading interpretations of data, there always remains some uncertainty about the strength of the effect sizes.

Chapter 5

A twin-case study of developmental number sense impairment
Chapter 5

A twin-case study of developmental number sense impairment

Abstract

The current study reports on 9-year-old monozygotic twin girls who fail to make any progress in learning basic mathematics in primary education. We tested the hypothesis that the twins’ core math problems were deficits in number sense that manifested as impairments in approximate and small number systems, resulting in impairment in non-symbolic as well as in symbolic processing. While age matched controls (eight typically developing girls) scored highly, the twins scored at chance on all number sense tasks. More specifically, on a non-symbolic comparison task, even in the simplest ratio condition of 1:2, and on a subitizing task including only numbers under 4, the twins performed at chance and significantly below the same age control group. Responsiveness to an intervention promoting number sense is discussed. As differences between verbal and performance IQ suggest, there seems to be a high degree of specificity in the twins’ developmental number sense delays. The concomitant impairments for visual-spatial processing and working memory in the twins might explain the failure to develop number sense.

Published as:

Introduction

This study examines the profiles of two 9-year-old monozygotic twins characterized by a failure to make any progress in learning basic mathematics in primary education. After a global assessment of the twins’ general cognitive capacities and academic skills, we hypothesized that the twins suffered from a severe number sense deficit expressed in impairments in the approximation system as well as in the small number system, resulting in impairment in non-symbolic as well as in symbolic processing (Dehaene, 2011; Feigenson, Dehaene, & Spelke, 2004; Molko et al., 2003). Some consider those skills as core deficits in learning basic mathematics (Price, Holloway, Räsänen, Vesterinen, & Ansari, 2007). Secondly, we hypothesized that their responsiveness to common interventions aimed at improving number sense skills would be low because they lacked skills that were at the basis of math development (e.g., Wilson et al., 2006). Particularly visual-spatial and/or working memory impairments may be seen as part of the root cause of the failure to develop number sense (Dehaene, 2011; Geary et al., 2009). Delays in higher-level math skills (i.e., solving problems) despite normal achievement in other academic skills may be the result of deficits in elementary numerical processing (Price et al., 2007). According to recent insights, elementary numerical processing encompasses two numerical systems: A small number system with exact representations for numbers under four and an approximate number system (also called the Approximate Number System-ANS) which enables young children to approximately compare magnitudes (Dehaene, 2011; Feigenson et al., 2004; Molko et al., 2003). The numerical ratio is the signature of the approximate number system (Halberda & Feigenson, 2008; Izard, Sann, Spelke, & Streri, 2009): Babies can correctly distinguish two pairs of non-symbolic magnitudes with a 1:2 ratio (Xu, Spelke, & Goddard, 2005) and three-year olds with a 3:4 ratio (e.g., 6 and 8 dots). Six-year olds can compare magnitudes with a 5:6 ratio, and this increases to a ratio of 10:11 in adults (Halberda & Feigenson, 2008). Nine-month-old infants seem to notice that 5+5 is not 5 but 10, because these comparisons are based on the easiest ratio of 1:2 (McCrink & Wynn, 2004). Wynn (1992) and Dehaene (2011) demonstrated that, next to approximate representations, infants also have exact representations for numbers, but this is limited to numbers under four. Typically developing children know for instance that 1+1 equals 2 and not 1 or 3, referred
to as subitizing (Wynn, 1992). The early capacity to track small numbers of objects may be conditional to inferring around the age of 3 or 4 that any set, however large, must have a precise number (Dehaene, 2011).

Neuroimaging studies reveal a specific area within the intraparietal lobes, the IntraParietal Sulcus (IPS), to be an important neural region for representing and processing quantity. This area in the brain responds to all the modalities of number presentation, and its activation varies according to whether the numbers are small or large, close or distant (e.g., Dehaene, 2011, Molko et al., 2003; Price et al., 2007). The neural region for processing quantity is tightly linked to the regions for space and time in the parietal brain area (Assel, Landry, Swank, Smith, & Steelman, 2003; Bull, Espy, & Wiebe, 2008; Dehaene, Spelke, Pinel, Stanescu, & Tsivkin, 1999). A recent study by Gebuis and Reynvoet (2012) demonstrated that comparing non-symbolic magnitudes is not a pure numerical process, but also involves weighting different visual parameters such as stimulus diameter, surface area, and density. They give the daily life example of choosing a train compartment: You will choose the one that looks least dense, without estimating the number of people. In the same vein, it is suggested that relations among magnitudes may arise from the strength of correlations between number and space in the natural environment (Cantlon, 2012). In line with these outcomes, fMRI studies revealed that approximation activates the IPS, but that activation also leaks to brain areas that code for location, size, and time similar to activation caused by visual-spatial tasks (Dehaene, 2011; Dehaene et al., 1999). Later on in development when the mental number line refines, influence of visual-spatial skills also becomes evident on a behavioral level (Dehaene, 2011): When representing numbers on a number line ranging for instance from 0-10 (Laski & Siegler, 2007), mental visual-spatial representations are essential for understanding that the number 9 should be put diametrically opposite 1 and the number 5 in the middle.

Likewise, working memory skills may be closely related to numerical development (Dehaene, 2011; Geary et al., 2009), as for instance, Wynn’s subitizing tasks with numbers under four demand that the child holds in mind pieces of information presented at different times. Five-month-old infants see one or more toys that are then hidden behind a screen. The babies have to keep this information in mind and compare this to how many toys they see the experimenter add or remove before they can make a decision about whether the presented solution to a sum is correct or incorrect.

We propose that apart from impairments in learning and retrieving arithmetic facts or deficits in executive functions (Wilson & Dehaene, 2007), children with serious delays in learning math may show serious number sense deficits and fail to compare magnitudes (e.g., choosing the largest amount) even when numbers are below 4 (e.g., Von Aster & Shalev, 2007). In those children not only processing of symbolic, but also non-symbolic magnitudes may be affected (Feigenson et al., 2004; Price et al., 2007), probably in combination with visual-spatial and working memory impairments (Geary et al., 2009; Gebuis & Reynvoet, 2012).

If serious math delays relate to a number sense deficit, training basic number sense skills such as comparing amounts might be a way to remediate these math problems. Previous research targeting children with developmental dyscalculia (DD), demonstrated for instance that number line representations of 8-10 year olds can improve by means of an intervention in which children practice locating the correct spot on a number line (Kucian et al., 2011). The only interventions that proved successful in the remediation of dyscalculia though, included children who were able to process magnitudes, but who did so slower than their peers (e.g., Kucian et al., 2011; Moeller, Neuberger, Kaufmann, Landerl, & Nuerk, 2009). If the twins indeed demonstrate an inability to compare numbers under four, they may show low responsiveness to an intervention that trains the ability to compare numbers.

Case Report
Nine-year-old monozygotic twin girls were referred to the university clinic for a psycho-educational and neurological examination because of a serious delay in math development and the absence of evidence for response to intervention. Despite a daily individual instructional program and daily individual tutoring during the first two-and-half years of primary education, all attempts to teach the twins basic math skills such as counting, estimating, comparing numbers, and simple addition and subtraction sums up to 10 had failed. The twins continued making errors in counting, and their ability to solve simple sums under ten and memorizing those problems varied per day.

The twins were born very pre-term (Gestational Age: 27 weeks and 5 days) and had a Very Low Birth Weight of 980 gram (N) and 960 gram (J), respectively. Survivors with a VLBW or very pre-term birth (Gestational Age (GA) ≤ 33 weeks) are at a substantially greater risk for developmental disorders than children with a higher birth weight, higher gestational age, or term-born children (Aarnoudse-
Moens, Weisglas-Kuperus, Van Goudoever, & Oosterlaan, 2009; Taylor, Espy, & Anderson, 2009). According to a meta-analysis on neurobehavioral outcomes of VLBW children and/or children born very pre-term (Aarnoudse-Moens et al., 2009), these children perform significantly poorer on mathematics, reading, and spelling compared to term-born peers. Combined effect sizes were -0.60 SD for mathematics, -0.48 SD for reading, and -0.76 SD for spelling. Moreover, pre-term birth may intrude on cortical development and the development of brain connectivity or myelination (Sansavini, Guarini, & Caselli, 2011). FMRI data (Isaacs, Edmonds, Lucas, & Gadian, 2001) showed reduced grey matter in the left intraparietal sulcus (IPS) in a preterm group with calculation difficulties at age 15, while a preterm group without calculation problems did not show these brain abnormalities. In addition to grey matter abnormalities, Clark and Woodward (2010) found white matter abnormalities in preterm born children with poor visual-spatial working memory performances at age six. In this study, the white and grey matter abnormalities were demonstrated at term equivalence age long before cognitive problems emerged.

This study started with a general cognitive assessment followed by assessment of the twins’ math, reading and spelling skills. Subsequently, numerical processing skills were analyzed in a step-wise design, starting with higher-order math skills and ending with the most elementary number sense skills. Data on standardized tests were contrasted with the mean score of the reference group as reported in the accompanying manuals. In so far as standardized tests were not available, the data of the twins were contrasted with data of eight typically developing control children using a variant of the t-test adapted to the comparison of a single participant’s performance to that of a small control sample in single case studies (Crawford & Howell, 1998; Crawford, Howell, & Garthwaite, 1998).

**General Abilities**

**Intelligence.** The twins’ intelligence profile as measured with the WISC-III NL 3rd edition (Wechsler, 2005) was extremely disharmonious; N and J’s performance IQ was 61 and 72 respectively which is substantially below the norm, whereas their verbal IQ was within one standard deviation, 87 and 90 respectively. In Table 5.1 an overview of their standardized scores is presented. Their scores on Block Design, Picture Arrangement, Mazes, and Object Assembly were 2 standard deviations or more below the norm. On the subtest Arithmetic, they performed very poor: 3 standard deviations below the norm. Scores on the subtests Similarities, Coding, Vocabulary, and Comprehension were within the normal range. Working memory was poorly developed according to their low score on the WISC-III subtest Digit Span (cf. Barsky & Siegel, 1992). Assessed with a standardized word span task instead of the digit span test (Lehide Diagnostische Test [Leiden Diagnostic Test]; Schroots & Van Alphen de Veer, 1979), N scored more than 3 standard deviations below the norm (IQ Equivalent = 55) but J scored just outside the normal range (IQ Equivalent = 82).

**Table 5.1**

*Standardized scores on general cognitive abilities*

<table>
<thead>
<tr>
<th>Subtest</th>
<th>Standardized scores N</th>
<th>Standardized scores J</th>
</tr>
</thead>
<tbody>
<tr>
<td>Block design¹</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Mazes¹</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Picture arrangement¹</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Object assembly³</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Arithmetic¹</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Similarities¹</td>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td>Coding¹</td>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>Vocabulary¹</td>
<td>12</td>
<td>8</td>
</tr>
<tr>
<td>Comprehension¹</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>Digit span²</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Verbal IQ²</td>
<td>87</td>
<td>90</td>
</tr>
<tr>
<td>Performance IQ²</td>
<td>61</td>
<td>72</td>
</tr>
<tr>
<td>Word span²</td>
<td>55</td>
<td>82</td>
</tr>
</tbody>
</table>

¹ (M = 10; SD = 3). ² (M = 100; SD = 15)

**Math skills.** The Tempo-Test-Rekenen [rate-based test for math] (de Vos, 2001) with a total of 200 addition, subtraction, multiplication, and division problems of increasing difficulty up to 100 was administered. The test starts with single digit problems and ends with more complex multi-digit problems. Within 3 minutes, children need to solve as many sums as possible. The score is the total of correct sums minus the incorrect ones. Both girls scored among the lowest 10% compared to the age norm group. It was striking that both girls never used a counting strategy (e.g., finger counting) or any other manifest procedure to solve the sums.
Reading and spelling skills. Since Grade 3, the twins had some problems with spelling when words followed complex rules. On a standardized Dutch dictation task, ‘PI-dictee’ (Geelhoed & Reitsma, 2004) they both scored among the lowest 10% compared to the age norm group, mainly due to errors in complex rules that often included the concepts ‘short’ and ‘long.’ An example of a complex rule in Dutch is the short and long vowel: In ‘pap’ [cereal] ‘a’ is pronounced short while ‘a’ in ‘wagen’ [car] is pronounced long.

A standardized word reading task, ‘Eén-Minuut-Test’ [one minute test], was administered to assess word reading skills. Within one minute, words have to be read aloud from a list. The standardized score is based on the number of words read accurately, the child’s gender, and age (Brus & Voeten, 1973). N and J both scored more than 2 standard deviations above average (standardized scores are 19 and 17 respectively). On the pseudo-word reading test ‘Klepel’, that assesses how many pseudo-words can be read accurately within 2 minutes (Van den Bos, Lutje Spelberg, Scheepstra, & de Vries, 1994), the twins’ standardized scores were 15 (N) and 16 (J), which was more than one standard deviation above the norm.

In summary, the twins’ deficit provides us with a rare opportunity to explore the nature of a serious failure to learn math. Combined with failure to make any progress in math, we expected that the twins might have an inability in processing symbolic as well as non-symbolic magnitudes. They might lack any sense of the meaning of magnitudes, represented by deficits in the approximate as well as the small number system, probably as a result of serious visual-spatial and working memory impairments in accordance with their intelligence profiles.

Experimental Studies of Number Sense

The first step in the experimental studies was to test the twins’ representations of numerical magnitudes under 10. Estimating the locations of different specific numbers from all parts of the numerical range under 10 is assumed to be a higher-level math skill on which typically developing kindergartners perform well (Siegler & Ramani, 2008). As a critical test the twins located numbers on a number line with 0 on one end and 10 at the other with no other numbers or marks in-between. In this study, eight age-matched typically developing girls participated as a control group (mean age = 10.4, SD = 3.5 months).

Numerical Magnitude Representations

Material and procedure. Numerical magnitude representations (Laski & Siegler, 2007) were assessed with a computerized version of the Number Line Estimation Task, programmed in E-Prime 2.0; see panel I in Figure 5.1. Children marked the numbers 1, 2, 3, 4, 5, 6, 7, 8, and 9, each presented twice in random order, on a line with 0 and 10 printed on the ends (for a detailed description see Siegler & Ramani, 2008; Laski & Siegler, 2007). All estimates were automatically saved to two decimal places. Based on the ratings, each child’s linear fit (expressed in $R^2$) was computed. $R^{2\log}$ represents a logarithmic fit for the data of the 0-10 number line, commonly providing the best fit in preschool-age: Children’s estimates exaggerate differences in magnitudes of smaller numbers and compress differences in larger numbers (Siegler & Ramani, 2008; Laski & Siegler, 2007).

A transition to a linear fit occurs during preschool and kindergarten. $R^{2\text{linear}} = 1$, if all numbers are marked exactly on the corresponding spot.

Results. The 8 control girls scored almost perfectly linear: $R^{2\text{linear}} = .99$ for the 0-10 number line. The scores of N and J, both $R^{2\text{linear}} = .72$, were significantly less linear ($t(7) = 25.46, p < .001$ (N and J)). A paired sample t-test revealed that the linear fit of the control group was significantly higher than the logarithmic fit of .89 [$t(7) = 25.44, p < .001, d = 8.66$]. The difference between the logarithmic [$R^{2\log} = .79$ (N) and $R^{2\log} = .76$ (J)] and linear fit (both $R^{2\text{linear}} = .72$) of the twins was smaller compared to the difference observed in the control group [$t(7) = 14.97, p < .001, d = -35.69$ (N) and $t(7) = 16.81, p < .001, d = -43.80$ (J)]. These results indicated that the twins had not developed a linear representation for number symbols up to 10, but relied more on a logarithmic representation for numbers below 10, which also becomes evident from Figure 5.2. They were able to say the numbers in correct order but did not understand the rank order of the numbers’ magnitudes. Both estimated under 5 rather well but underestimated numbers beyond 5.

The twins’ poor visual-spatial skills may explain their poor performance on the number line task (Dehaene 2011; Geary et al., 2009). A certain spatial ordering of quantities is necessary when mentally placing numbers on a line: One needs to know for instance that the number 9 should be put diametrically opposite 1 and the number 5 in the middle.

In line with the finding that the meaning of symbols for numbers up to 10 was not obvious to them, we hypothesized that the twins might be unable to understand, approximate and manipulate non-symbolic numbers, and as a
result not connect number symbols to quantities. As a critical test, we tested the hypothesis proposed in the literature that they might have an inability to compare number magnitudes (Halberda & Feigenson, 2008). Secondly, we tested how they would approach a non-symbolic magnitude comparison task, when magnitudes were not processed automatically. Moeller et al. (2009) used an eye-tracker in a dot counting task to test the use of a counting strategy in two boys with dyscalculia. They hypothesized that the two 10-year-old boys with dyscalculia in their study, would show a stronger increase in reaction time (RT) and number of fixations as more dots had to be counted. Compared to the control group, the two boys indeed showed a much steeper regression slope in RT and number of fixations, suggesting that they needed to count at least in some proportion of the trials, whereas the control group processed more automatically, resulting in faster RT’s and less fixations.

Numerical Magnitude Comparisons
In line with Moeller et al. (2009), we administered a numerical magnitude comparison task using an eye tracker (Tobii T120), to find out whether the twins...
were sensitive to the difficulty of the number comparison tasks. We expected that with increasing ratio (ranging from 1:2 to 7:8), the controls would respond slower and demonstrate more eye-fixations to compare the two sets of dots (Moeller et al., 2009). The twins might show a similar but slower pattern or, in case they were unable to solve such problems, a completely aberrant pattern.

Material and procedure. In this numerical magnitude comparison task children were asked to select the largest set of dots ranging from 1-16 (n = 96). Presented ratios were 1:2 (n = 48), 3:4 (n = 24), 5:6 (n = 12), and 7:8 (n = 12). In each ratio, all possible dot pairs were presented, therefore n varied among ratios. In presenting sets of dots, surface area, dot diameter, and density varied (cf. Halberda & Feigenson, 2008). Dot color of the left set was orange and of the right set pink. Background color of the slides was blue; see panel II in Figure 5.1. There was no time limit. After the child gave a verbal response (orange or pink), the experimenter pressed the spacebar. Next a smiley appeared (for 1 second) in the center of the screen whereupon the next trial appeared. An eye-tracker was used to make manifest how the twins’ approach differed from the control group’s approach. The influence of increasing ratio was evaluated by performance (percentage correct), RT (mean duration of eye-fixations), and number of fixations, using Tobii Studio version 2.2.6 (Tobii Technology, 2010). Based on the ratio condition, regression slopes for RT and number of fixations were calculated for each control child and compared to the slopes of the twins (see Moeller et al., 2009).

Results. Descriptive statistics for percentage correct, RT, and number of fixations, split by ratio are displayed in Table 5.2. The control group had a high mean score (96.26% correct). In line with previous findings (e.g., Halberda & Feigenson, 2008), the ratio 7:8 was most difficult for the control group (on average 91.67% correct; see Table 5.2). The twins scored significantly lower than the control group: Overall percentage correct of N and J was at chance and equaled 39.18% and 49.48%, respectively [t(7) = -20.39, p < .001, d = -21.62 (N) and t(7) = -16.71, p < .001, d = -17.72 (J)]. Also when split by ratio, percentage correct of the twins was at chance in each ratio and significantly lower than in the control group; N for ratio 1:2: t(7) = -1045.58, p < .001, d = -1109.00 and J for ratio 1:2: t(7) = -928.67, p < .001, d = -985.00; N for ratio 3:4: t(7) = -603.68, p < .001, d = -640.30 and J for ratio 3:4: t(7) = -447.17, p < .001, d = -474.30; N for ratio 5:6: t(7) = -560.12, p < .001, d = -549.10 and J for ratio 5:6: t(7) = -480.93, p < .001, d = -510.10; N for ratio 7:8: t(7) = -654.78, p < .001, d = -694.50 and J for ratio 7:8: t(7) = -392.37, p < .001, d = -416.17.
The control group had increasing RT's as well as increasing eye-fixations when the ratio became more difficult (see Table 5.2), with the longest mean reaction time and highest number of fixations in the ratio 7:8 condition. The twins' RT's and eye-fixations by contrast did not increase with ratio (see Table 5.2), resulting in a flatter slope compared to the controls, that is, falling outside the range of the box plots (Figure 5.3). Findings thus suggest that the twins tried to solve the problems in the most difficult ratio the same way as in the easiest condition. This is striking as the more difficult ratios, with a smaller difference between numerical magnitudes, take longer to process and more fixations are needed before deciding which dot pair is largest. A shorter reaction time and lower number of eye fixations in combination with the low performance across all ratios [1:2, 3:4, 5:6, 7:8] strongly suggest that the twins failed to implement a strategy to compare amounts and were just guessing. The control children by contrast did have more fixations and a longer fixation duration in the more difficult ratios, as their RT's and eye-fixations increased with ratio. Even though there was no time limit, controls did not count the dots but seemed to compare surfaces covered with dots, the dots' density, and the dots' diameter by moving the eyes between the two sets of dots.

In summary, these findings indicate that the twins were unable to compare non-symbolic magnitudes under 16 (Halberda & Feigenson, 2008), even within the easiest ratio of 1:2. These outcomes thus suggest that the twins' approximate number system was underdeveloped (Dehaene, 2011; Halberda & Feigenson, 2008). Maybe due to underdeveloped visual-spatial skills, they were unable to weight different visual parameters (i.e., dot size, surface area) and therefore might be unable to estimate differences between two sets of dots (Gebuis & Reynvoet, 2012). The question arises whether next to the approximate number system, the number system was underdeveloped (Dehaene, 2011; Halberda & Feigenson, 2008). However, in the subitizing task, where objects were covered, added or removed, we could not rule out that poor working memory skills might have interfered with solving the addition and subtraction problems (Dehaene, 2011). In the number line and magnitude comparison tasks, demands on working memory were lower, because stimuli remained present.

The twins' poor visual-spatial skills may have had an impact, because the bears remained in the same place and the twins had to keep a vivid and realistic image of the objects hidden behind the screen: a kind of mental photography (Dehaene, 2011). Likewise, we argued that visual-spatial skills may have affected the twins' numerical processing skills in the number line and magnitude comparison task (Dehaene, 2011; Gebuis & Reynvoet, 2012). To test seriousness of visual-spatial problems in solving number sense tasks, we used subtests of the magnitude discrimination tests designed by Siegel (1971) with a decreasing visual-spatial component. The first test tapped strongest on visual-spatial abilities: A child must recognize simple size differences by choosing the largest of two solid areas. The ability to recognize equivalent sets of dots is less dependent on visual-spatial skills than magnitude comparison but more so when visual-spatial relationships between dots remain the same. All three tests were programmed in MS PowerPoint (see panels IV, V, and VI in Figure 5.1). In the for 1.5 seconds, bears are added or subtracted. During 1 second the child sees bears fly out/in. After 1.5 seconds, the screen is removed (duration 1 second). The result (1 to 3 bears) is then displayed without a time limit and the child has to say whether the displayed answer is correct or not (too many or too few bears). They were not asked to provide the exact answer (e.g., there should be 2 bears).

**Results.** The twins scored at chance level [38% and 57% correct for N and J], which was significantly below the mean of the control group scoring 94.26% correct \( t(7) = -5.85, p < .001, d = -6.20 \) (N), and \( t(7) = -4.19, p < .01, d = -4.44 \) (J)). There were no response biases, meaning that the presented outcome (either correct or incorrect) and type of problem (addition or subtraction) did not affect the results. Outcomes thus imply that the twins did not know that one bear plus another bear equals two, and not three or four bears (Wynn, 1992), which indicates that they missed exact representations for numbers up to three.

To summarize, the twins had a strong and stable deficit in comparing amounts (up to 16), even when quantities were under four. Results of the subitizing task indicated that the twins' small number system was underdeveloped. However, in the subitizing task, where objects were covered, added or removed, we could not rule out that poor working memory skills might have interfered with solving the addition and subtraction problems (Dehaene, 2011). In the number line and magnitude comparison tasks, demands on working memory were lower, because stimuli remained present.

The twins' poor visual-spatial skills may have had an impact, because the bears remained in the same place and the twins had to keep a vivid and realistic image of the objects hidden behind the screen: a kind of mental photography (Dehaene, 2011). Likewise, we argued that visual-spatial skills may have affected the twins' numerical processing skills in the number line and magnitude comparison task (Dehaene, 2011; Gebuis & Reynvoet, 2012). To test seriousness of visual-spatial problems in solving number sense tasks, we used subtests of the magnitude discrimination tests designed by Siegel (1971) with a decreasing visual-spatial component. The test tapped strongest on visual-spatial abilities: A child must recognize simple size differences by choosing the largest of two solid areas. The ability to recognize equivalent sets of dots is less dependent on visual-spatial skills than magnitude comparison but more so when visual-spatial relationships between dots remain the same. All three tests were programmed in MS PowerPoint (see panels IV, V, and VI in Figure 5.1). In the
control group, subtests IV and V were terminated when the first 10 consecutive items were correct.

**Magnitude Comparisons with Differential Influence of Visual-Spatial Skills**

**Material and procedure.** In the first subtest (panel IV in Figure 5.1), the largest of two solid area figures (for instance two rectangles or two circles, differing in size) had to be chosen (80 items in total). This task tested whether the twins were able to compare size of objects. Comparing objects or figures differing in size (e.g., big/small) is basic to ordering and therefore assumed to be an early developing numeracy skill (Torbeyns et al., 2002), typically tapping on visual-spatial abilities. In the second subtest of the magnitude discrimination test (equivalence) a dot pair was presented at the top of the computer screen and at the bottom, the child had to select the same amount of dots out of four dot pairs (80 items); Figure 5.1, panel V. Magnitudes ranged from 1-9. Subtest three in Siegel’s battery (conservation) was the same as subtest two, with the exception that the same amount of dots presented at the bottom was differently arranged than at the top of the screen (20 items); Figure 5.1, panel VI. Thus in comparison to subtest two, children could not rely on the visual arrangement of dots in this subtest, but instead numerical processing was necessary.

**Results.** The twins scored significantly below the mean of the control group (see Table 5.3 for mean scores) on magnitude discrimination of solid area figures \( t(7) = -51.15, p < .001 \) (N), and \( t(7) = -31.25, p < .001 \) (J), which highlights the seriousness of visual-spatial delays. On the equivalence and conservation subtest, the twins scored at chance as well and significantly below the mean of the control group (see Table 5.3 for mean scores): Equivalence \( t(7) = -47.50, p < .001, d = -47.50 \) (N), and \( t(7) = -48.75, p < .001, d = -48.75 \) (J), and conservation: \( t(7) = -4.49, p < .01, d = -4.49 \) (N), and \( t(7) = -4.76, p < .01, d = -5.05 \) (J).

In other words, there was no evidence for modulation of number sense scores as a function of the visual-spatial task component. The twins consistently scored at floor and the level of visual similarity did not modulate their performance. The overall low performance, even on the simplest magnitude discrimination task, strengthens the hypothesis that the twins’ poor visual-spatial skills might be at the root of delays in normal number sense.

If the visual-spatial component of the equivalence and conservation subtests was simplified by only presenting numbers under four, the twins scored higher. On equivalence, N and J scored on 27 items that met this criterion 70% and 67% correct, respectively. Similarly, on conservation both scored 80% correct on 5 items under 4. In comparison to the control group, the twins performed significantly better on items with amounts under four, compared to items with amounts of four and higher in both tests \( t(7) = 112.32, p < .001, d = 1907.07 \) (N equivalence) and \( t(7) = 104.31, p < .001, d = 1645.44 \) (J equivalence), and \( t(7) = 95.72, p < .001, d = -1386.34 \) (N conservation) and \( t(7) = 95.58, p < .001, d = -1382.43 \) (J conservation)]. This finding may indicate that the twins were to some extent sensitive to the complexity of the visual-spatial component of the task.

<table>
<thead>
<tr>
<th>Test</th>
<th>Number of items</th>
<th>Scores N</th>
<th>Scores J</th>
<th>Scores controls (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mental arithmetic (raw score)</td>
<td>200*</td>
<td>13</td>
<td>8</td>
<td>27.2 (reference group from manual)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td><strong>R^2</strong> = .72</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td><strong>R^2</strong> = .72</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td><strong>R^2</strong> = .79</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td><strong>R^2</strong> = .76</td>
</tr>
<tr>
<td>Number line 0-10</td>
<td>18</td>
<td>R^2lin = .72</td>
<td>R^2log = .79</td>
<td>R^2lin = .72</td>
</tr>
<tr>
<td>Subitizing (% correct)</td>
<td>37</td>
<td>38</td>
<td>54</td>
<td>94.26 (9.07)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td><strong>R^2</strong> = .99</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td><strong>R^2</strong> = .89</td>
</tr>
<tr>
<td>Solid figures (% correct)</td>
<td>80</td>
<td>48.75</td>
<td>68.75</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0)</td>
</tr>
<tr>
<td>Equivalence (% correct)</td>
<td>80</td>
<td>52.50</td>
<td>51.25</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0)</td>
</tr>
<tr>
<td>Conservation (% correct)</td>
<td>20</td>
<td>50%</td>
<td>45%</td>
<td>90.63 (9.04)</td>
</tr>
</tbody>
</table>

*maximum score. Score equals the number of correct answers within 3 minutes.

In summary, the findings on the number line task, the eye-tracking magnitude comparison task, the subitizing task, and the magnitude comparison tasks of Siegel indicate that the twins had a serious delay in number sense, in the approximate as well as the small number system (Dehaene, 2011; Feigenson et al., 2004; Molko et al., 2003). All number sense tasks were solved at chance level (Dehaene, 2011; Siegel, 1971; Wynn, 1992). There seemed to be a high degree of specificity in the twins’ developmental number sense delays as differences between verbal and performance IQ suggest. At the root of the twins’ number processing impairments there might be poorly developed visual-spatial skills.
A twin-case study of developmental number sense impairment

(Assel, Landry, Swank, Smith, & Steelman, 2003; Bull, Espy, & Wiebe, 2008; Dehaene, 2011; Geary et al., 2009; Gebuis & Reynvoet, 2012) as well as working memory delays. All number sense tasks included demands on both visual-spatial skills and working memory albeit not to the same extent (Geary et al., 2009). For instance, the demands on working memory might have been stronger in the subitizing task as in this task elements were covered, whereas in the other number sense tasks, elements remained visible (Dehaene, 2011).

Outcomes support the hypothesis that deficits in both numerical systems, the approximate and exact small number systems, can be core problems when pupils have serious math problems (Price et al., 2007). Yet we cannot rule out the possibility that impairments in visual-spatial skills and working memory caused underdevelopment of number sense and that persistence of those impairments interfered with improvement during the years in school.

The question remains whether in such cases core number sense skills can be trained. Previous research showed positive effects of number line training in children with dyscalculia (Kucian et al., 2011; Moeller et al., 2009). However, contrary to the twins in this study, children in these studies had some basic numerical understanding but processed magnitudes slower than typically developing children due to the need to use a counting strategy (e.g., Kucian et al., 2011; Moeller et al., 2009). An intervention was implemented to find out whether the twins were responsive to practicing comparing numbers.

**Responses to Intervention**

**Number Race.** In the second half of Grade 3, both girls received, twice weekly, a computer-assisted intervention (Räsänen, Salminen, Wilson, Aunio, & Dehaene, 2009; Wilson et al., 2006; http://sourceforge.net/projects/numberrace/) for about three months (N 23 times, J 16 times) under supervision of an undergraduate student. Although there is no consensus about the effectiveness of math intervention programs, the NR game especially aims to train assessment of quantities, the counting routine, the link between symbols and quantities, and the understanding that number and space are linked (Dehaene, 2011; Räsänen et al., 2009). As number comparison skills seem to be a core deficit in case of the twins, this program was chosen. J started the intervention four weeks later than N (resulting in 23 sessions for N and 16 sessions for J).

In the intervention that aimed at comparing amounts and connecting amounts to symbols the twin girls were instructed to select the highest number from two options presented either as dots, numbers or as addition or subtraction problems (Räsänen et al., 2009; Wilson et al., 2006). Later tasks increased in difficulty due to the use of numbers or addition and subtraction problems. The rest of the week they received care-as-usual including remedial teaching.

During the whole intervention period, N was unable to decide which amount of dots was largest in the NR games, even for numbers below 4. Throughout all sessions, N mostly guessed with the exception of comparing 2 and 4. She thought that 2 dots were more than 4 and stuck to that idea. Attempts for demonstrating the concept of more and less (e.g., counting and then locating the numbers on the number line) were unsuccessful. At the end of the intervention, N’s notion of more and less had not improved despite suggestions of the supervisor to count and decide with the help of a number line showing all numbers between 0 and 10.

In the first two levels of the NR (14 levels in total), J often guessed which amount was larger and only counted the dots when encouraged by the experimenter. She benefitted, however, from counting and marking both amounts on the number line when choosing the larger amount. Without a number line, she guessed which of the two options was larger resulting in scores at chance level. Throughout the sessions (see Figure 5.4) her performance showed an upward tendency ($R^2 = .79; \beta = .45$) and she reached the highest level possible ($M level = 10; small addition problems up to 5$), different from N ($M level = 7; comparison

![Figure 5.4. Achieved levels of the Number Race game by session, for N (dashed line with triangle markers) and J (solid line with squares).](image-url)
of verbally spoken numbers up to 9). J outperformed N (t (23) = 3.14; p < .01; d = .53), probably not due to improved magnitude comparison skill but because she was more successful in using additional tools like the number line to solve the games.

**Discussion**

Are there core deficits that account for problems in acquiring arithmetic skills similar to a phonological deficit in reading? The current study reported on two monozygotic 9-year old twins born very preterm with a VLBW, who seem to have a remarkably underdeveloped approximate and small number system. Both girls fail to understand concepts like bigger/smaller even though they use these words. Most surprisingly, they failed to notice what most young children normally understand without any training, namely that one bear plus another bear equals two, and not three or four bears. They were unable to represent, discriminate, and operate on small numbers under four (Wynn, 1992). Unlike other examples of a number sense deficit described in the literature (e.g., Moeller et al., 2009), the twins were not just slower in number comparisons but completely failed, even within the easiest ratio condition of 1:2. The eye-tracking data revealed that the twins did not show the expected increase in RT and number of fixations by task complexity, whereas the control children did. Outcomes of the control children were in line with previous findings that magnitudes with a small numerical distance are more difficult to compare than magnitudes with a large numerical distance (numerical distance effect; e.g., Halberda & Feigenson, 2008). As the twins scored significantly below the mean on all ratios, outcomes strongly suggest they were unable to compare the sets of dots and were just guessing.

In summary, outcomes imply that both numerical systems, the approximate number system as well as the small number system, were underdeveloped (Dehaene, 2011; Feigenson et al., 2004; Molko et al., 2003), which may explain why further math development was hindered (Dehaene, 2011; Price et al., 2007). Their subitizing deficit in combination with their difficulties in comparing non-symbolic amounts (i.e., knowing what is larger) might explain why they failed on marking numbers on a number line. This typically taps into the ability to compare quantities and relate number symbols to quantities (Laski & Siegler, 2007). The twins fail in solving problems, probably because solving simple problems strongly builds on understanding quantity relations (Krajewski & Schneider, 2009). When learning sums like 2 + 3, a child needs to understand that the outcome must exceed 3.

Unlike other studies that report attempts to improve number sense skills, our attempts to enhance quantity comparison skills failed. N remained unable to compare amounts, even when encouraged by the experimenter to count correctly and to look up numbers on a number line. J failed as well in comparing the numbers but, unlike N, she did benefit from applying a counting strategy and using a number line when decisions had to be made about the largest amount. However, J never used these tools spontaneously, but only when she was encouraged and coached by the experimenter in applying these strategies. Thus only one twin benefited somewhat from an intensive training in comparing numbers and locating numbers on a line. Like the two boys with dyscalculia in the study of Moeller et al. (2009), the children with dyscalculia in the study of Kucian et al. (2011) performed slower on number sense tasks, whereas the twins in our case study did not solve number sense tasks beyond chance level.

**Visual Spatial and Working Memory Skills**

Results corroborate the hypothesis that visual-spatial skills and working memory (like scores on puzzles, object assembly and digit span of the WISC) commonly associated with dyscalculia (e.g., Geary et al., 2009), may be at the root of the twins’ number sense impairment. Comparing amounts of dots requires estimating the surface area, stimuli diameter and density, and thus visual-spatial skills (Dehaene et al., 1999; Gebuis & Reynvoet, 2012). As we discussed above, subitizing as assessed in a test modeled after Wynn’s task, but other number sense tasks as well, strongly demand working memory (Dehaene, 2011). The results indicate a high degree of specificity in the developmental number sense delay: The twins’ score on sub-tests of the WISC-III with a strong visual-spatial and working memory component was substantially below the norm, whereas their score on verbal IQ-tests was within the normal range. In the same vein, the Number Race intervention may not have been effective or only somewhat effective because the twins lacked visual-spatial and working memory skills that are indispensable to compare amounts – the main activity in the Number Race. Other types of training that target visual-spatial and working memory skills might have been more beneficial than the Number Race.

In summary, the current findings corroborate the hypothesis that an inability to weight visual parameters is at the root of impairments in the approximate and
small number system. Likewise, holding information in mind and manipulating that information, may on its own or combined with visual-spatial deficits also be at the root of the ability to solve number sense tasks (Dehaene, 2011; Geary et al., 2009). We speculate therefore that in these specific cases number sense may not normalize until their visual-spatial skills and working memory catch up. It may have consequences for treatment whether or not at the root of delays are visual-spatial and/or working memory deficits. Hence, more important than the putative category ‘dyscalculia’ is to specify underlying impairments.

**Future Directions**
The finding that the twins seem to manifest no processing difference in comparing magnitudes with different numerical distance, might point to a weakened parietal representation of non-symbolic magnitudes (Price et al., 2007). According to, for instance Wilson and Dehaene (2007) and Price et al. (2007), at least some children with serious math problems typically show less activation of the parietal cortex while comparing numbers. Likewise, the fMRI studies of Isaacs et al. (2001) and Clark and Woodward (2010) revealed that in children born preterm (in case of the twins at Gestational Age 27) and with calculation problems, grey matter may be reduced in the left intraparietal sulcus (IPS). The IPS is detectable in fetuses around GA (pregnancy week) 26-28. As the twins were born at GA 27, this may have affected prenatal brain development and the IPS in particular. There is also evidence from fMRI studies that during non-symbolic magnitude comparison, nonverbal visual-spatial cerebral networks are activated (Dehaene et al., 1999). As visual-spatial skills are undeveloped, atypical activation of nonverbal visual-spatial cerebral networks could be present in the twins (Clark and Woodward, 2010).

It is also possible that impaired neuronal networks might account for the twins’ severe number sense deficit (Kucian et al., 2006) as task performance is mostly dependent on the cooperation of diverse brain areas (Micheloyannis, Sakkalis, Vourkas, Stam, & Simos, 2005). A consolidated neuronal network is necessary when diverse brain areas are involved in task performance. Weaker activation of a network could be the result of non-reinforced synaptic connections due to lack of experience (Kucian et al., 2006), which is not very plausible here. Next to weak connections between brain areas, intraparietal disconnection after damage to a focal region of subcortical white matter might cause math problems (Rusconi et al., 2009; Sansavini et al., 2011). As a next step in analyzing the twins’ math problems it might make sense to test structural connectivity between brain areas.