TYPICAL AND ATYPICAL DEVELOPMENT OF MAGNITUDE PROCESSING

Fruzsina Soltész

Ph.D. Thesis

2010

Doctoral School of Psychology: Cognitive Development Programme

Faculty of Psychology and Education

Programme director: Prof. Magdolna Kalmár
Supervisor: Prof. Valéria Csépe
Co-mentor: Dr. Dénes Szűcs

President: Prof. Magdolna Kalmár
Readers: Prof. István Czigler
Dr. Attila Márkus

Committee: Dr. Anikó Kónya
Dr. Nóra Katona
Dr. Julianna Szendrei

Eötvös Loránd University of Sciences

Budapest

2010

Acknowledgements

I would like to take this opportunity to thank my supervisor Prof. **Valéria Csépe.** She provided intellectual and personal support and guidance throughout my PhD studies which helped me a lot to begin my research at the first place.

I owe a tremendous debt of gratitude to my co-mentor, Dr. **Dénes Szűcs**, for his invaluable support, patience and encouragement both in professional and private life without which this thesis would not be complete.

I am also very grateful to Prof. **Usha Goswami** at the Centre for Neuroscience and Education, University of Cambridge. She is a great boss and person to work with and I am really enjoying being a part of her intellectually stimulating and colourful research group.

I would also like to thank all the adult and children **volunteers** for participating in my experiments. The studies reported in this thesis were partly funded by University of Cambridge and data were collected, stored and analysed by using the facilities provided by the Centre for Neuroscience and

Thank you also to my fellow PhD students at the Centre for Neuroscience and Education in Cambridge: **Sonia White**, for her invaluable help in data collection, **Donna Bryce** and **Clare Killikelly** for the English proof-reading and corrections.

I would like to thank my friends, especially **Gayaneh Szenkovits** and **David Asfaha** for all the thoughtful conversations we had throughout these years. They always managed to bring the smile to my face at any times.

I also wish to thank my family, and especially to my mother **Erika Jankovics** for her emotional and financial support throughout these years; and also for her unwavering love and belief in me without which I could not get any close to where I am now.

Introduction

Numerical abilities are very important in our life. From its simplest forms of like comparing two sets of food based on the number of food items in each set, to the more complex forms of for example linear algebra and calculus all contributed to the survival and development of the human race. In the Western cultures arithmetic knowledge is crucial for the functioning of societies, and for the well being of the individuals as well. Arithmetic skills are indispensable for many of the professional qualifications and people with better arithmetic skills are usually well off in jobs and can lead a life of better quality, than their peers with weaker arithmetic abilities. For these reasons, the research on arithmetic abilities got popular among psychologists and neuroscientists in the past few decades. The development of arithmetic abilities and the disability of learning mathematics, in peculiar, are of accentuated interest because the charting of *typical and atypical development* of numerosity may shed light on the nature of cognitive functions behind numerical knowledge. By mapping the development of these cognitive functions, hopefully both the individual and the society will benefit: people in *education* and in *special education* will hopefully draw on the findings of *developmental cognitive neuroscience* when they develop and improve teaching and remediation methods.

In this thesis my aim is to introduce and discuss the literature of numerical cognition, with an emphasis on *typical and atypical development of arithmetic abilities*. Based on some unresolved issues and hot debates in the literature, an event-related potential (ERP; extended with principal component analysis [PCA]) and behavioural study will investigate the cognitive factors behind the atypical development of numerical abilities. Magnitude representation, among several other abilities (spatial- and body representations, working memory, IQ, executive functions) will be tested (developmental dyscalculia, chapter III).

Another behavioural study will look into the debated nature of the co-development and interdependency of magnitude representation, memory and counting knowledge in kindergarten children (chapter IV).

Two further studies will explore the development of magnitude representation: the interaction of numerical and physical magnitudes will be tested in an ERP and behavioural experiment, considering the role of attentional and executive functioning as well (chapter V; extended with an adult PCA study: chapter VI); and the speed and automaticity of accessing numerical representations will also be examined in children, with ERP and behavioural measures (chapter VII).

The results will be discussed in light of the importance of considering a broader range of basic cognitive abilities in the research of numerical abilities. Basic skills like executive functioning, working memory, and inhibition of prepotent and incorrect responses seem to be highly relevant in both typical and atypical numerical development. In my opinion, arithmetic abilities are by far more complex than it could be explained by one single cognitive module. This supposed module, the *core analogue magnitude representation* may provide a good basis for the understanding of relative and approximate magnitudes, but this is yet far away from the problems children have to cope with in school. Rather, the *network* of co-developmental routes and interactions of several basic and specific abilities have to be considered and examined in more detail during the research of numerical development and developmental dyscalculia.

I. Magnitude representation – is it a module?

Abstract

By most researchers in the field of numerical cognition, number representation is assumed to be domain specific, innate, neurally hardwired, and accessed in a fast and automatic fashion (for an overview, see Dehaene et al., 1998; or Cantlon et al., 2008). These attributes are exactly the proposed attributes of a cognitive module. In this chapter, I review the literature of magnitude representation in the light of cognitive modularity. For the organization of this chapter, I will use Coltheart's (1999) definition of cognitive module. Results of behavioural and neuro-cognitive, developmental and comparative studies will be introduced under the appropriate sections. Both pros and contras for the modular assumption of numerical processing are discussed. The empirical studies in the present thesis will build on some hypotheses, unresolved questions of this review chapter.

1. Cognitive and neuro-cognitive models of magnitude representation

The most widely accepted cognitive model of number representation is the so-called 'triple code model' (Dehaene and Cohen, 1995; Dehaene et al., 1998). In this model, number representation is considered to function as an evolutionary inherited, highly specialized cognitive module. Other models[1] of number representation (e.g. McCloskey, 1992; Campbell and Clark, 1988: complex coding model; Noel and Seron, 1992: preffered code model) are not explicitly claimed to be modular. However, these models could be also viewed as highly domain-specific and modular in practice. *Domain-specific*, because neither of these models are stated or assumed to be valid for other domains of cognition. For example, no one intends to explain language processing

[1] Because the cognitive models of number representation and processing have been already introduced in great detail elsewhere (see Szűcs, 2003; Soltész, 2004), I will not discuss them in the present thesis.

by a cognitive model of magnitude representation, or vice versa. And *modular*, because the content and function of these models are described as separate 'boxes of cognition'.

The 'triple code model' consists of three brain circuits which are responsible for the three different functions contributing to number representation. The first system is called the 'analogue magnitude representation' or the 'the core quantity system', hosted in the horizontal intraparietal sulcus (brain imaging (Dehaene and Cohen, 1997; Dehaene et al., 1999; Eger et al., 2003; Pinel et al., 2001; Simon et al., 2002; for a review, see Dehaene et al., 2003) and lesion studies (e.g. Cipolotti et al., 1991)). Approximation and comparison of numerical magnitudes are supported by this core quantity system. The analogue magnitude representation is complemented by two other systems, if the situation requires. The left gyrus angularis and the perisylvian areas host the verbal system which is responsible for calculations executed in verbal form (e.g. overlearnt multiplications, verbal problems). The verbal system also contributes to exact calculations (Spelke and Tsivkin, 2001). The bilateral posterior superior parietal areas play a role in the orientation of spatial attention, indispensable for instance in number comparison, subtraction, or during reading and writing numbers with more place values (Grossberg and Repin, 2003).

1.1. The analogue magnitude representation

The analogue representation of numerical magnitudes can be best described by a so-called 'mental number line' (Restle, 1970). This mental number line can be imagined as a *mental* instantiation of a real number line, except that the values and boundaries among the different magnitudes are noisier than on the real number line (Sekuler and Mierkiewitz, 1977; Moyer and Bayer, 1976). Noisiness means here that a certain value is not represented as a discrete, individual point on this mental number line; rather, all values have a distribution around a central point, which central point represents the given value in an *approximate* manner. The representational distribution of the adjacent values overlap; the farther the two values are from each other, the less the overlap between their representational distributions, so forth the easier it is to discriminate them. Further, the farer we 'look' on this mental number line, the less accurately we 'see' the represented values. Representations of numbers at the beginning of the 'number line' (e.g. 1-10) are less noisier or overlapping than numbers at the farer left side of this number line (large numbers). This noisiness of the representation leads to some very specific predictions regarding the performance during numerical judgements: the closer the two magnitudes are to each other, the larger their representational overlap, the more difficult it is to discriminate between them – so forth the number of errors and response time should increase.

1.1.1. Markers of the magnitude representation

Comparison or discrimination of numerical magnitudes obey Weber's law, similarly to all other perceptual variables: the discrimination performance is the function of the *ratio* of the to-be-compared values. The most widely used behavioural marker of the analogue magnitude representation is this *ratio effect*, which incorporates the so-called *numerical distance effect* (Moyer and Landauer, 1967) and the so called *number size effect*. The numerical distance effect (see Figure I/1) reflects the phenomenon that it is more difficult to discriminate two numbers which are close to each other than two numbers which are further apart from each other. For example, we are faster and more confident to say that 2 is smaller than 10, than to say that 9 is smaller than 10. The same phenomenon can be observed during the perceptual comparison of non-numerical analogue magnitudes. For example, in case of luminance, physical size or colour, it takes more time and effort to discriminate values which are closer to each other than to discriminate values which are farther apart.

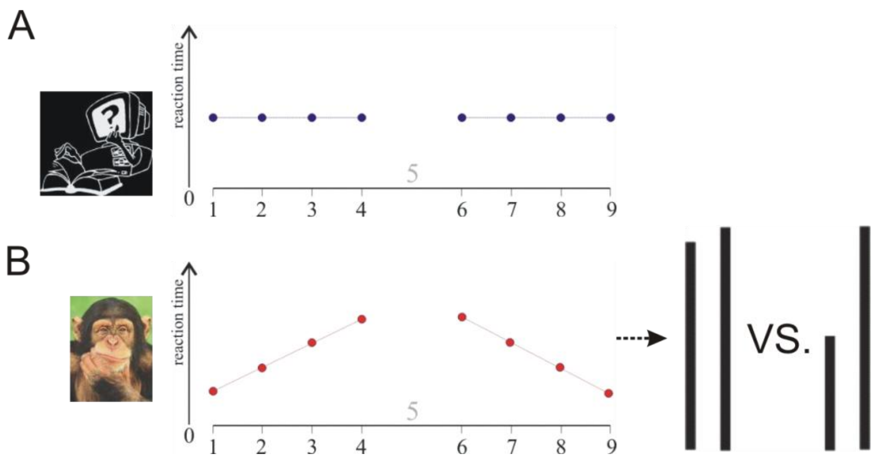


Figure I/1: Illustration of the distance effect. A: hypothetically, if numerical values were discrete and non-overlapping with each other, it would take the equal amount of time to say “bigger than 5” both to 6 and to 9 – just as computers do. B: However, with animals, including humans, this is not the case. The closer the target number to the reference, the longer the decision time. Discrimination of numerical magnitudes follow a similar pattern to that of physical magnitudes, e.g. it is easier to find the longer stick when the difference between the two sticks is larger, than when the difference between the sticks is smaller.

The *numerical size (or problem size) effect* inherently contributes to the *ratio effect.* The larger the values are, the larger distance is required for the same level of discriminability. For example, in case of let’s say 1 and 2, the NDE is 1; just as it is in case of 7 and 8. However, the discrimination of 7 and 8 is more difficult, because their ratio (7:8) is much smaller than the ratio of 1 and 2 (which is 1:2).

The *ratio effect* in behaviour has consistently been found in adults performing approximate and exact number comparison; in children performing exact and approximate number comparison; and in animals performing approximate number comparison (for reviews see: Gallistel and Gelman, 2000; Feigenson et al., 2004; Dehaene et al, 2004; Cantlon et al., 2008). The numerical distance effect or the ratio effect has also been demonstrated in blood-oxygen-level-dependent (BOLD) responses in functional magnetic resonance (fMRI) studies (Piazza et al., 2004, 2007) and also in stimulus-related electrical activity measured by electro-encephalography (EEG). In EEG, the distance effect has been demonstrated in adults in experiments using visual stimuli (for example: Grune, 1993; Dehaene, 1996); using acoustic stimuli (Szűcs and Csépe, 2004, 2005). Also in children using symbolic (Arabic numerals) and with non-symbolic (dot collections) stimuli (Szűcs, Soltész et al., 2007; Temple and Posner, 1998) and in adolescents with developmental dyscalculia, using symbolic stimuli (Soltész et al., 2007). EEG signatures of the numerical distance effect have also been shown in 3- and 7-month-old infants (Libertus et al., 2008; Izard et al., 2008).

As the above list of experimental numerical distance and ratio effects also suggests, the representation of numerical magnitude is independent of modality; Arabic numerals, written number words, spoken number words, acoustic tones, visual patterns are all transformed to a notation-independent and abstract representation of magnitude (Libertus et al., 2007; Brannon, 2003; Barth et al., 2003; Pinel et al., 2001).

To sum up, the analogue magnitude representation, hosted in the horizontal intraparietal sulcus, is supposed to provide the basis for numerical cognition. This analogue representation is proposed to be an evolutionary inherited *modul* in the animal kingdom. Further, it is also assumed to play the key role in higher calculation abilities in humans.

2. Properties of a cognitive module

The term *modularity* refers to the concept (or to the principle) that cognition can be split into a collection of small, independent and specialized subprocesses (Marr, 1982). These subprocesses, or *modules* are considered to be isolated from each other and supposed to be highly specialized to one task.

The concept of modularity is central to cognitive science. In his milestone book, Fodor (1983) further elaborated the attributes and properties of a cognitive module. However, although widely used, Fodor's definition of the cognitive module is mostly misinterpreted according to Coltheart (1999). Fodor's description of a cognitive module does not imply that the possession of one or the other property were a necessary condition for a system to be a module. Rather, according to Coltheart, neither of the following attributes is necessary or sufficient for the definition of a cognitive module; rather, these are only properties, and not crucial conditions of a cognitive module. These attributes, or properties, are the following:

(1) Domain specificity

(2) Innateness

(3) Informationally encapsulated

(4) Hardwired (specific neural networks are responsible for the function)

(5) Not assembled (from other cognitive subcomponents)

(6) Automatic, and

(7) Fast (in terms of processing and speed of access)

2.1. Domain specificity

Following the terminology of neuropsychology, a function is considered to be a functionally independent module, if it shows *double dissociation* from other functions*:* it can be impaired independently of other functions, and vice versa, other functions can be impaired leaving our function of interest intact.

Neuropsychological cases proved the existence of double dissociations of numerical abilities. Brain-damaged patients who suffered focal parietal lesions exhibited severe deficits in counting and calculation, while their performance in other high level cognitive tasks is intact (for example language, semantics) (Cipollotti et al., 1991; Dehaene and Cohen, 1997). There is example for the opposite pattern as well: severe semantic or language deficits can be accompanied with intact numerical and calculation abilities (Dehaene and Cohen, 1995).

A very similar criterion like the double dissociation is applied for the diagnosis of *developmental dyscalculia* as well: children who show severe deficits in arithmetic and mathematics are diagnosed with developmental dyscalculia only if their other cognitive abilities (e.g. reading, spelling, IQ) are in the normal range. Although the diagnosis criteria of developmental dyscalculia do not yet bear a wide consensus, it is generally agreed upon that at least two years discrepancy has to be found between the child's actual performance and his age-matched peers' performance, while their other cognitive abilities are supposed to be intact. It is a widely accepted working hypothesis in the research of developmental dyscalculia that the deficiency of the core system of analogue magnitude representation is the reason behind this learning difficulty (Butterworth et al., 1999; Feigenson et al., 2004).

2.1.1. Or rather domain generality?

When we talk about numbers or space, we use the same adjectives: 'small' and 'big'. When we compare different types of magnitudes, we can ask the same questions: 'Which one is more?' or 'which one is larger?'. In fact, all the comparisons across different magnitudes and modalities yield the ratio- or distance effects, indicating highly similar cognitive mechanisms, or indicating only one, general mechanism behind these comparisons (for a review see Cohen Kadosh et al., 2008). It has been already proposed that number is represented by a domain-general magnitude system (Moyer and Landauer, 1967; Gallistel and Gelman, 2000). This domain-general magnitude system, obeying Weber's law, is in function whenever one-dimensional, analogue magnitudes have to be compared or manipulated. For example, the processing of space and time has been proved to overlap and interfere with the processing of numbers, leading some researchers to construct a new "theory of magnitude" (ATOM, Walsh, 2003).

Regarding *space* and number, the most important behavioural marker of the numerical magnitude representation, the ratio effect has been found when the lengths of lines had to be compared (e.g. Fias et al., 2003; Cohen Cadosh et al., 2008). Sensitivity to the changes in surface area (i.e. space) also showed the ratio effect in children (Brannon et al., 2006). That is, for example, when Elmo faces of varying sizes were shown to 6-month-old infants, reacted (dishabituated) to change in surface, when the ratio of the to-be-compared surfaces reached 1:2, but they failed to detect the change at the ratio of 2:3. Also, performance in some numerical tasks reflects strong associations with space and orientation. One demonstration of a strong connection between space and time is the spatial-numerical association of response codes (SNARC) effect (Dehaene, Bossini and Giraux, 1993)**.** When subjects are asked to classify a number as even or odd, they responded faster with left hand when the numbers in question were small (e.g. 1 or 2), and they responded faster with their right when the numbers were large (8 and 9). The SNARC effect is due to the spatial nature of the representation, rather than to some associations with the hands: when hands were crossed, so that the left hand was on the right side of the body, and vice versa, the left hand (on the right side) reacted faster to large, the right hand (on the left side) reacted faster to the small numbers (Dehaene et al., 1993). Another demonstration of the relationship of numbers and space is the attention bias effect (Fischer et al., 2003): small numbers orient attention towards the left, large numbers orient attention towards the right in space. This bias was measured by reaction times when a target had to be detected as soon as possible either on the left or in the right side of the screen, after the presentation of a one digit number in the middle on the screen. Reaction was faster for targets on the left side after a small number was shown, while responses were faster for a target on the right if a large number was shown. A further example for the space-number association is the line bisection effect (Calabria and Rossetti, 2005): when asked to midpoint a line constructed of 'two', subjects tended to misplace the midpoint towards the left; while when the line was made up of the word 'nine', their responses were biased towards the right. And as for a last example for the association of space and time, patients with spatial neglect also 'neglected' one side (most commonly the left) of the mental number line: when asked to name the approximate middle point between the numbers 3 and 12 (Zorzi et al., 2002), patients tend to pick a number from the right side of the mental number line, e.g. 9.

Regarding *time* and number, Gibbon, Meck and Church (1983) have already shown in rats that the estimation of number and the estimation of time followed identical patterns and were influenced in the same manner by drugs (Church and Meck, 1984). In rats again, the representations of space and number is supposed to overlap because they were able to calculating rate via combining information from number and time (Gallistel and Gelman, 1992). Of course there are some experiments with humans as well. In humans, the execution of mental arithmetic was disrupted by a secondary time estimation task, but not by a rotor tracking or by a visual detection task (Brown, 1997). Although the author identified executive functioning and attentional resource distribution behind these findings, Walsh (2003) claims that it is rather the same mechanism, which is involved in both number and time tasks, which was the source of this interference. In 6-month-old (VanMarle and Wynn, 2006) and in 10-month-old infants (Brannon et al., 2008) time has recently also been shown to be processed in a similar way than numerical magnitudes. Just like adults, these infants detected changes in time (as measured by event-related potentials) in function of the ratio of the change. Older children's (3-8 years-olds) performance in time comparison tasks is also found to be subject to Weber's law (Droit-Volet, Clement and Wearden, 2001).

Besides space and time, comparison performance across several physical magnitudes shows the very same effects of distance and size. The discrimination efficiency of angles (Fias et al., 2003), the mental comparison of sizes of animals (Moyer, 1973), or even the comparison of the pitch of sounds (Rusconi et al., 2005; 2006) are subjects to Weber's law, letting us to infer that a common representation, or a common mechanism operates behind these discrimination processes. Whether the overlap among these dimensions occur at the level of mental representations, or at the level of procedures executed during approximation, comparison, decision making and response organization, cannot be answered on the basis of behavioural data. Behavioural data are not sufficient for the extrication of these cognitive processes; data from brain imaging and electrophysiological studies will provide further insight into the localization of these processes in space (brain) and in time (milliseconds).

2.2. Innateness

Evidence for the innate nature of the representation of numerical magnitudes emanate from comparative studies including animals and human infants. It is reasonable to assume that a construct is evolutionary inherited, if this it bears the same functional and anatomical properties across different species, and is also present in human infants. Accordingly, comparison and approximation of numerical magnitudes are not unique to human adults or even to the human species. Animals present some kind of "number sense" (Dehaene, 1997). For instance, rats are able to transfer numerical information across perceptual modalities (Gibbon et al, 1984), or monkeys are able to generalize newly trained rules across completely different objects and situations. Furthermore, animals show the same signature of magnitude representation: the ratio effect. For example, in a non-symbolic addition task, where human and monkey participants had to choose the roughly corresponding set of dots after the addition of two sets, a very similar pattern of the effect of ratio was found (Huntley-Fenner, 2001; Cantlon and Brannon, 2007) (Figure I/2).

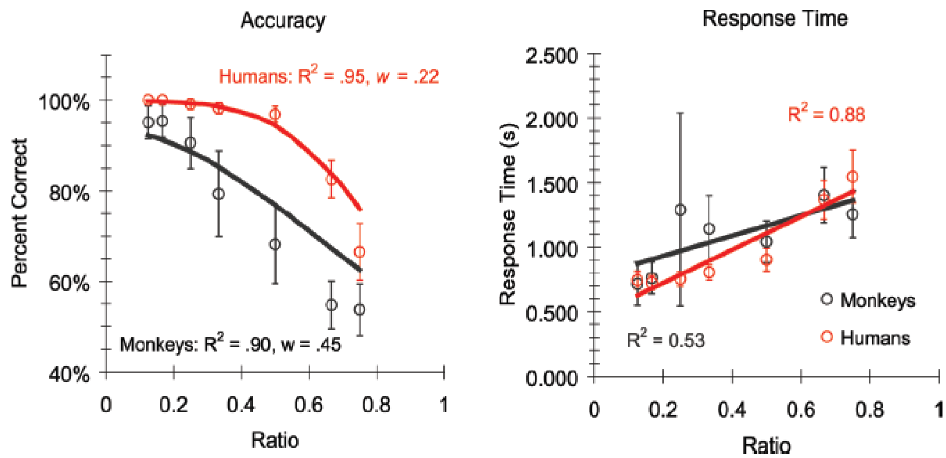


Figure I/2: Figure taken from Cantlon and Brannon (2007). Both in accuracy and response time, monkeys and graduate students exhibited a very similar pattern of ratio effect.

In sum, approximation and comparison performance of animals are indeed show similarities with that of humans (e.g. Mechner, 1958; Platt and Johnson, 1971; Whalen et al., 1999; Gallistel and Gelman, 2000; Jordan and Brannon, 2006a,b), suggesting that we share an evolutionary inherited, common representation of numerical magnitude.

2.2.1. Numerical abilities of human infants

Preverbal human infants possess a similar sense of numbers than animals do. Infants are able to discriminate different sets of objects based on their numerosity, as it has been found in several habituation and looking preference studies with infants. In a *habituation paradigm* babies are shown displays of different objects (e.g. two dots) until they lose their interest in the particular display. This decrement of interest can be measured by the looking time. When infants are shown some habituation trials in sequence, their looking time drops from display to display. After reaching habituation, which is let's say a 50% drop in looking time, the display is changed (test trial); if infants spend significantly more time with looking at it, than they did the previous habituation display, we can say that they dishabituated to the particular test display. An other way to investigate infants' abilities is the *looking preference paradigm*. In this paradigm, two objects are shown at the same time; if infants ‘prefer’ or ‘surprised by’ (i.e. look longer at) one than the other object, one can infer that they differentiated between the two objects in some ways. An alternative of the looking preference paradigm is the *violation of expectation:* babies look longer at things which were unexpected for them, so forth violated their expectations.

Using a habituation paradigm Starkey and Cooper (1980) were the first to show that 4-7 months old babies were able to discriminate 2 dots from 3 dots. This study was followed by several others, all replicating and strengthening the finding on the numerical abilities of infants (/static visual stimuli/: Antell and Keating (1983), Strauss and Curtis (1981), Starkey, Spelke and Gelman (1990); /auditory stimuli/: Bijeljac-Babic, Lipton and Spelke (2003); /moving, grouped stimuli/: van Loosbroek and Smithsman (1990), Wynn, Bloom and Chiang, (2002); /intermodal, auditory and visual matching/ Starkey, Spelke and Gelman, (1990)). The ratio effect has also been found: infants could discriminate large sets with a ratio of 1:2, but they fail with a ratio of 2:3 (Xu, 2003; Xu and Spelke, 2000, Xu, Spelke and Goddard, 2005; Rouselle and Noel, 2008; Brannon, Abbot and Lutz, 2004, Lipton and Spelke, 2003; Wood and Spelke, 2005; Xu, 2003). The ratio-sensitivity of children of course develops with age: at 9-10 months-old, they discriminate sets with the ratio of 2:3; adults cope with approximately 7:8 (Xu and Arriaga, 2007).

Beyond magnitude comparison, infants are capable of performing simple calculations (Koechlin, Dehaene and Mehler, 1997, Simon, Hespos and Rochat, 1995, Uller, Huntley-Fenner, Carey and Clatt 1999). Wynn (1992), using the looking preference paradigm, demonstrated that 4-5 month-old infants were ‘surprised’, their expectances were violated, by erroneous outcomes of simple additions. In the experiment, Wynn placed one puppet in front of the baby, and then covered it with a curtain. She placed another puppet behind the curtain, still in front of the infant’s eyes. Finally, the curtain was removed, revealing either one, or two puppets. When the incorrect outcome was shown (1+1=1), infants looked longer – they were surprised. This experiment was extended to show that infants are able to perform not only additions, but simple subtractions as well (Wynn, 1992). And it has been also proven that infants recognize the *exact* change in number, instead of the *direction* of change:

they looked longer at instances of 1+1=3 than 1+1=2 as well. If infants perceived the direction of change only, i.e. 'something else was put there, so there should be *more*', then the 3 dolls as an outcome would have been just as correct as the 2 dolls (Wynn, 1995).

2.2.2. Number comparison experiments with infants: Methodological issues

The most important methodological challenge in the above introduced number comparison experiments is that perceptual variables are correlated with number. These variables are interdependent both of each other and of numerosity and it is impossible to control for all of them at the same time. For instance, if intensive properties (individual item properties, like item size) are kept equal in a particular trial, extensive properties (properties of the set, like summed surface of all items in a group) will inevitably co-vary with number, and vice versa. With a simple example, a collection of 6 apples is not only more, but physically also larger than a collection of 3 apples. In nature, 'more' usually correlates with 'bigger' (number of individuals in a group, number of pieces of food, etc.). Infants may have well relied on these simple perceptual features of the sets, instead of the more abstract numerosity. Namely, several of the early studies did not control for these perceptual correlates of the stimuli (Starkey and Cooper, 1980; Antell and Keating, 1983; Strauss and Curtis, 1981, 1984; Starkey et al., 1990; van Loosbroek and Smitsman, 1990; Cooper, 1984; Huntley-Fenner and Cannon, 2000; Wynn, 1992; Simon et al., 1995), making infants' proposed numerical performance indistinguishable from their perceptual performance. In fact, it was found that if sum of the occupied surface area is rigidly controlled for, children are not able to discriminate collections of sets based on their numerosity. Rather, infants represent quantities based on continuous overall amount (Mix, Huttenlocker and Levine, 2002, Rouselle et al, 2004). Further, 6-8 months old children were dishabituated to change in circumference but not to number, where the overall circumference across the sets of dots was kept constant (Clearfield and Mix, 1999).

Others, for example Xu and Spelke (2000) tried to control for co-varying perceptual variables. During habituation trials, the summary of the surface of the dots were varied and only numerosity was kept constant. In the test trials, when the number of the dots changed, the sum surface of the dots was picked from the range of the habituation trials, so forth was not new for the infant. However, the sum circumference of the dots cannot be controlled at the same time as the sum surface, because surface and circumference are both the functions of the radius (r) of the circle (surface: $r^2\pi$; circumference: $2r\pi$). So forth circumference did indeed correlate with numerical change, and could have well explained Xu and Spelke's results (see also: Mix et al., 2002).

These methodological issues, among others, will be relevant and will be discussed in more detail in Chapter IV.

2.2.3. The role of enculturation and language in numerical development

Adult humans are clearly beyond the approximate comparison of the number of sets of objects, or adding up two dolls. Let alone mathematics, physics and science, adults of the Western culture use and manipulate exact numbers and numerical symbols in their everyday life. Due to enculturation, our concepts of numbers change radically and take a quantitatively different form from that of preverbal infants and nonhuman animals. Magnitudes are not handled as approximate values anymore; they are mapped with exact and discrete values, which enable us to carry out more complex calculations than magnitude comparison. The labels or symbols we use to denote numerical magnitudes let us to know *exactly* of how many we are talking about. With language, we *digitalized*, marked out mental number lines. There is no noise, no overlap among these values any more, allowing us to calculate that for example 6+7 equals to *exactly* 13, and not *approximately* some more than the fingers we have on our hands.

Language and numerical enculturation is inevitable for the formation of the concept of exact numerical magnitudes (e.g. Dowker, 1997). For example, native Amazonian people do not use exact number words above four: 'pug'=one; 'xep xep'=two; 'ebapug'=three; 'ebadipdip'=four; 'pug pogbi'=one hand; and above these magnitudes, they discriminate roughly between some and many. These people discriminate among 1, 2, 3 and four relatively exactly. But above four, they use their categories, for example they say “some” to 7 and 8, and they say “many” to 14 and 15 (Gordon, 2004; Pica et al., 2004).

In an experiment conducted by Spelke and Tsivkin (2001), Russian-English bilinguals were taught some exact calculation problems (e.g. 16×8=128) and some approximate calculation problems (e.g. 16×8= ~120,130). In the test phase, both the exact and approximate problems were akked, in both languages. They performed better on exact problems when they had to solve it in the language they were taught in, while language did not matter with the approximate problems. In the approximate problems their performance was independent from the language in which these problems were taught to them.

Language is doubtlessly necessary for the development of exact numbers and calculation (both phylo- and ontogenetically). What is still a matter of debate in the literature of numerical development is whether the analogue magnitude representation can be specialized to the abstract concept of number without language, or whether language is inevitable for the refinement of the analogue magnitude representation. The importance of this distinction has a huge impact on what we think is important for children’s numerical development. On the long run, it also affects the teaching methods of mathematics, not to mention the ruling theories behind the remediation of mathematics learning difficulties.

In Chapter IV, I will focus on this debate and test the role of verbal abilities, together with other general cognitive abilities, in the development of magnitude and number representation.

2.3. Informationally encapsulated

This is the one presumed property of a cognitive module which can be instantly and doubtlessly confuted. Numerical and non-numerical information interact with each other, as demonstrated with the *size congruity effect* (Banks and Flora, 1977; Besner and Coltheart, 1979; Henik and Tzelgov, 1982). This effect was found in a *numerical Stroop paradigm* (for a review see Szűcs et al., 2007).

Analogous to the classical Stroop paradigm (Stroop, 1935), in the numerical Stroop paradigm subjects have to make decisions based on the perceptual features of the stimuli, while the conceptual meaning conveyed by the stimuli is not relevant. In the classical colour-word Stroop paradigm, it is the ink colour of the word in which it is written what is relevant for the task, while the meaning represented by the word is irrelevant. Simple as it seems, participants still have some difficulties in this task. Indeed, the meaning of the word still influences subjects' answers: for instance they slow down and make more mistakes when the word 'red' is written in green ink colour. In the numerical Stroop paradigm, the physical sizes and numerical sizes of the two simultaneously presented stimuli are varied, and participants have to indicate the physically larger stimulus independently of their numerical meaning. Again, although a simple perceptual comparison would be sufficient to solve the task, numerical meaning intrudes the comparison of physical magnitudes and influences one's decision. This interference of numerical and physical magnitudes is called the size congruity effect. The *size congruity effect* can be delivered in the following three ways. In the congruent condition, numerical size difference between the two numerosities matches the physical size difference, e.g. 2 9. In the incongruent condition, the two dimensions (physical size and numerical size) are in conflict with each other, e.g. 2 9. In the neutral condition the irrelevant dimension is kept constant, e.g. 2 2. Cost in response time (RT) and accuracy in the incongruent condition compared to the neutral condition is called 'interference effect'. Gain in RT or accuracy in the congruent condition compared to the neutral condition is called 'facilitation effect' (Besner and Coltheart, 1979). In adults, it is consistently found that numerical meaning is automatically processed and interferes with performance, even though it is completely irrelevant for the task (Henik and Tzelgov, 1982; Tzelgov et al., 1992; Kaufmann et al., 2005; Rubinstein and Henik, 2005; Szűcs and Soltész, 2007).

With the initial version of this paradigm (Besner and Coltheart, 1979), one can examine whether physical size interferes with numerical magnitude, or not. When first applied, the numerical Stroop paradigm was theoretically the opposite of the original Stroop paradigm: participants were asked to decide upon the numerical size (the meaning of the symbol) and to ignore the physical size (the physical attribute of the symbol). In fact, physical size interferes with the numerical magnitude: participants slow down and commit more errors, when the physical dimension of the stimuli is incongruent with the numerical meaning (see Table I/1 for an illustration of congruency effects in both numerical and physical tasks).

	Neutral	Congruent	InCongruent
Number Task	2 7	2 7	2 7
Physical Task	2 2	2 7	2 7

Table I/1: Congruency effects in the physical and numerical versions of the number Stroop paradigm. Underscores index the correct choice. Neutral: the task irrelevant dimension is neutralized; Congruent: the two dimensions would lead to the same response; Incongruent: the two dimensions would lead to different responses.

Furthermore, luminance (darkness-lightness) is also found to interfere with numerical magnitude (Cohen Kadosh and Henik, 2006): it took longer for subjects to process big numbers when they were light and to small numbers when they were dark.

Taking all this evidence together we can conclude that the processing of numerical magnitudes is not encapsulated; other continuous magnitudes, like size and luminance, interfere with our decisions about numbers. However, these congruity effects do not necessarily imply that the interference among these different types of magnitudes originate at the level of representations. Interference can arise at later stages of processing, for example during decision, or response execution. Also, the interference among these dimensions could arise from the way we learn and talk about these dimensions: for instance most of the languages use the same words to express ‘small’ and ‘big’ along the dimensions of numerical and physical magnitudes, providing the same measure and the same responses across these magnitudes.

In Chapter V, the interaction between numerical and physical magnitude will be explored both with behavioural and electrophysiological measurements, in order to reveal the nature of these interference both in adults and young children.

2.4. Hardwired, domain-specific neural representation

The idea of a domain-specific brain representation of numbers postulates that there is a unique region in the brain, which responds exclusively to numerical stimuli, while it is insensitive to other types of stimuli. Also, this representation is assumed to be abstract and independent of the input notation (e.g. Arabic numerals, number words, or dots) (Ansari, 2008).

As mentioned before, the horizontal part of the intraparietal sulcus (IPS) is the proposed host of numerical representations. According to the 'Triple code model' (Dehaene and Cohen, 1995), the IPS is responsible for the coding of numerical magnitudes. Several neuroimaging studies have found number-related activity in this area of the cortex (for an overview see Dehaene et al., 2004), and there is no more than 5-7 millimetre discrepancy in the localization of number-specific effects in these studies (Dehaene et al., 2003). The IPS was found to show number-specific activity to symbolic (number words and Arabic digits, Pinel et al., 2004) and to non-symbolic stimuli as well (dots, Temple and Posner, 1998) (see Figure I/3). Furthermore, the IPS showed number-specific activity when neither comparison, nor any other arithmetic manipulation was required: the participants only had to detect a target number, a target letter or a target colour (Eger et al., 2003). The IPS was active only when a target number had to be found, and not when a target letter or colour. Conceptual manipulation with numbers, when magnitude comparison is not involved (e.g. identifying months with numbers) also activates the IPS (Cappelletti et al., 2009). Naccache and Dehaene (2001a) showed that the activity of the IPS was modulated by numerical distance, even without the conscious recognition of the stimuli (unconscious repetition priming). In this experiment, participants were shown number words with such a sort presentation time which made conscious recognition of these words impossible. After this priming stimulus, an Arabic digit appeared and participants were asked to decide whether this digit was smaller or larger than 5. The activity measured in the IPS was modulated by the priming stimuli: when the same number was presented by the prime (e.g. 'NINE') and by the target (e.g. '9'), the activity of the HIPS was found to be significantly lower in comparison to when the two numbers were not the same. Besides proving that numbers can be processed without conscious detection, this experiment also supports the notation-independent and abstract nature of number representation, because prime (number word) and target (Arabic number) were presented in different formats.

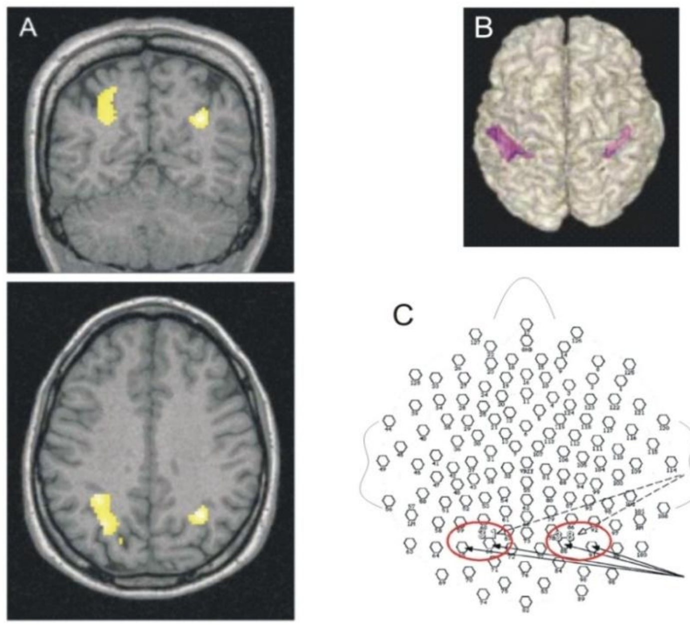


Figure I/3: Number-specific activity in the IPS (fMRI – A: from Eger et al., 2003; B: Pinel et al., 2004) and over the parietal electrodes (ERP – C: Temple and Posner, 1998).

The IPS showed number-specific activity during simple, numerical magnitude comparison tasks (e.g. Dehaene et al., 2003) and during more complex tasks, which require exact calculations (e.g. Simon et al., 2002).

Nieder et al. (2002; see also Nieder and Miller, 2004) demonstrated the existence of single neurons, which are sensitive to numerical distance. Nieder et al. examined the brain responses of monkeys, with intracranial single-cell electrodes directly placed into the brain[2]. They found that some cells (15% in the IPS) were tuned to one specific number (within the range of 1-4) and showed the highest activity when the preferred number was presented to the monkey. Also, the activity of these cells was somewhat noisy and showed certain distribution around the preferred value. These cells showed the highest sensitivity to their preferred number, however, also reacted to adjacent

[2] Intracranial recordings: the electrode(s) are led through the skull (cranium in Latin), so forth contacting the brain surface, or an assembly of neurons, or a single cell. Nieder et al. measured at the single cell level.

numbers as well. This property of number-specific cells is in accordance with the notion that the representation of numbers is noisy and the representation of adjacent magnitudes overlap – resulting in the distance effect during number comparison and number discrimination.

Although the fact that the activity of IPS can be manipulated by numerical tasks, it does not necessarily reflect number-specific functioning. The modulation of activity could be explained by other factors as well. For example, task difficulty could also result in the modulation of the activity of the IPS: small distance conditions can be considered as ‘difficult’, and large distance conditions can be considered as ‘easy’ conditions. Menon et al. (2000) examined the effects of task difficulty in order to disentangle numerical effects from task difficulty effects. The authors manipulated task difficulty by the number of operands and by the rate of presentation. Both manipulations exerted significant effects on the inferior frontal areas, but not on the IPS; which latter, in turn, was modulated by the numerical distance effect.

Furthermore, the parietal cortex is responsible for several different functions other than numerical processing as well. For example spatial and attentional functions and the coordination of the eyes and hands (see for example Culham and Kanwisher, 2001) are also in the functional repertoire of the IPS. Response selection and response execution also modulates the activity of the IPS, independently of number (Göbel et al., 2004). In order to disentangle these functions, some studies tried to separate numerical processing from eye-hand coordination, and from response selection and execution. Simon et al. (2002) showed that a segment of the IPS was sensitive exclusively to numbers, while other segments showed activity in function of spatial attention, hand- and eye coordination. Another way to separate these functions in the IPS, is to show whether it is modulated by the numerical distance in when response is not required. Using the adaptation paradigm, when participants were neither asked to pay attention nor required to give a response to numbers, the IPS was found to be modulated by numerical distance (Piazza et al., 2004; Ansari et al., 2006), independently of notation (Piazza et al., 2007).

Developmental fMRI studies have also shown that not only the pattern of behavioural responses, but the activation of brain areas are also highly similar in children and adults. For example, Kawashima et al. (2004) found no significant differences between 4-5 years-old children and adults in the activated brain areas during mental calculations like addition, and multiplication. Cantlon et al. (2006) contrasted the activation of IPS during shape and number adaptation in an adaptation paradigm. They found that the IPS in 4-year-old children was modulated in a similar way to that of adults. Other researchers showed that besides the specific activation of the parietal areas, there is a specialization going on in the brain during development. The prefrontal cortex, the hippocampus and the dorsal basal ganglia is more active in younger children than in their older peers, suggesting the more intensive involvement of attentional and working memory processes during the judgement of the outcomes of arithmetic operations (Rivera et al., 2005), non-symbolic (Ansari and Dhital, 2006) and symbolic number comparison (Ansari et al., 2005), and during number Stroop task (Kaufmann et al., 2006). Specialization of the proposed number-specific areas occurs while the role of executive functions and attention decline.

Regarding *developmental dyscalculia*, some researchers claim that the deficit of this number-specific neuronal 'module' of the IPS leads to difficulties with arithmetic and mathematics (Ansari, 2008; Hubbard et al., 2008; Price et al., 2007; Molko et al., 2003; Rotzer et al., 2008).

➡ In Chapter III, the proposed number-specific nature of DD will be investigated, via the examination of the behavioural and event-related potential markers of magnitude representation in adolescents with DD.

2.4.1. Not number specific, distributed representations

Contrary to the above number-specific assumption about the neural representation of numerical magnitudes, growing evidence suggests that the mental representation of numbers relies on a distributed network subserving the representation of other continuous magnitudes as well; including space, luminance, and time.

A thoughtful study of Shuman and Kanwisher (2004) confutes the supposed number-specific nature of the IPS. In a series of experiments, they (1) contrasted the activity recorded during number and colour comparison of dots (control for the experiment of Eger et al., 2003); (2) they explored subjects' brain responses during unconscious priming (following Naccache and Dehaene, 2001a), and; (3) they measured and contrasted the distance effect during number and colour comparison (control for the experiment of Pinel et al., 2001). Contrary to the referred studies, Shuman and Kanwisher used non-symbolic notation throughout the experiments. The authors could not replicate any of the IPS specific results of the previous studies. Furthermore, their results can not be criticized with being null findings[3], because the experimental manipulations exerted significant effects on the activity of IPS – which were not specific to numbers, but for example showed the same distance-modulated activity to colours and showed adaptation to repeated shapes. The authors emphasize several methodological issues concerning the previous studies. They claim that many of the previous studies did not control accurately for task difficulty, which modulates the IPS (e.g. numerical processing was contrasted which much simpler tasks, like fixation, colour detection etc). Also, they argue, that in the Naccache and Dehaene study (2001a) simple response priming could account for the results. Further, according to these authors, several previous conclusions were indeed based on null findings: IPS activation in numerical tasks was found to be significant compared to control, while non-numerical tasks did not differ significantly from the control task – so that, the argument is built on a non-significant finding (e.g. Simon et al., 2002; Pinel et al., 2004). Also, even in a well-controlled study (Eger et al., 2003) finding that some other domains do not activate the IPS automatically does not mean that nothing else would modulate the activity of the IPS. They conclude that there are two possible explanations for the contradicting results. As one alternative explanation, they propose that the IPS does not contain a neural instantiation of a domain-specific mechanism for numerical magnitudes. Rather, areas in the IPS are doubtlessly involved in number processing, but in several other processing

[3] Null finding or null effect is when an experimental manipulation does not yield any significant effects. However, a non-effect like this can not serve as evidence for the non-existence of the examined phenomena. The fact that we do not see something does not mean that it is not there; there are several other possible reasons for not finding it. A wrong measurement, a mistake in the experimental paradigm, a weak statistical power due to small and variable sample and to a small effect, among others, can lead to null effects. The way of reporting and interpreting such findings is under debate (e.g. Aberson, 2002); however, in psychology and cognitive neuroscience, a null effect is not considered to support the null hypothesis (which states that there is no effect) (Howitt and Cramer, 2007).

as well. As for another explanation, there might be some functional specialization within the IPS, in the range of neuronal populations, which are either too small, or are intermingled with each other in a way that (current) imaging techniques are not able to disentangle them.

In concert with the notion of an overlapping and distributed representation of magnitudes, Pinel et al. (2004) demonstrated that comparison of numerical magnitudes, physical sizes and luminance activated partially overlapping networks. Fias et al. (2003) also identified parietal regions equally responsible for line length, angle aperture and number comparisons, or even for ordinal judgements of the positions of the letters of the alphabet (Gevers et al., 2003).

Recent single-cell studies (Tuduscius and Nieder, 2007) found cells in the parietal cortex of primates sensitive exclusively to line length, other sensitive exclusively to number, while a certain amount (approx. 20%) responded to both line length and number. These results again support a distributed and partially overlapping network of magnitude representations.

To sum up, the analogue magnitude representation is best characterized as a broad domain, including a common representation and/or common measure mechanisms for many continuous variables, like size, luminance and number. The neural network subserving these functions is not modular; shared and overlapping networks lie behind these representations.

2.5. Assembled or not? – Cognitive components of numerical cognition

Whether the numerical representation is assembled or not, is a highly theoretical question. If we take the number-sensitive single neurons as a starting point for our argument, there is clearly no way, at least at the present state of science, to break this representation down into further particles.

But if we consider the cognitive levels of numerical cognition, the picture gets somewhat more complicated. Numerical abilities of humans, like set estimation and number comparison are presumably supported by two underlying core systems (Feigenson et al., 2004; Ansari and Karmiloff-Smith, 2002). One system is responsible for the fast and accurate enumeration of small sets, below 4-5 objects. This process of fast enumeration of small sets is called ‘subitizing’ (Mandler and Shebo, 1982; Trick and Pylyshyn, 1994). According to the object-file theory, during subitizing an object file is opened and stored for each individual item that appears in the visual field (Uller et al., 1999; Carey and Xu, 2001; Simon, 1997). As these object files represent discrete and individual objects independent of their appearance (size, shape, etc.), they form an ‘abstract’ representation in the sense that they store an abstract property of the set independent of its physical properties. As such, this ‘subitizing’ mechanism may provide the basis for the later abstraction of discrete numbers and preempts the understanding of the concept of natural numbers, where discrete intervals and a successor function of [+1] has to be applied (Ansari and Karmiloff-Smith, 2002). In this framework, the other core system is the analogue magnitude representation, which operates on larger magnitudes, and obeys Weber’s law. These two systems complement each other in the following way: the latter provides the sense of magnitudes, while the first makes exact enumeration and numerical abstraction possible.

An other possible distinction of the underlying systems of numerical cognition can be drawn between approximate and exact calculation. In approximate calculation, the analogue number representation and its presumed neural substrate, the IPS is at work; while exact calculation relies on verbal functions (Spelke and Tsivkin, 2001; Pica et al., 2004; Gordon, 2004), subserved by distinct language circuits (Dehaene and Cohen, 2005; Venkatraman et al., 2006). Further proof for the approximate-exact distinction comes from neuropsychological studies which found that brain damaged patients showed the dissociation of these two systems. Lesion in the right IPS caused the disruption of approximation, while exact calculation abilities remained intact (Dehaene and Cohen, 1991); there is evidence for the reverse as well: following an injury to the right parietal areas (including language areas), the patient had no difficulties with number comparison and approximation, while had severe impairments in exact calculation (Dehaene and Cohen, 1995).

2.6. Automaticity

Do we access numerical representations in an automatic fashion? Prior trying to answer this question, the term *automaticity* has to be clarified. Posner (1978) described automaticity as a process which is similar to reflexive behaviour, in that it runs without intention, attention or awareness. Although the role of attention is debated (e.g. Carr, 1992), it is agreed upon that an automatic process does not need monitoring to be executed (Zbrodoff and Logan, 1986).

In the literature of numerical processing, there are four main trends of showing that the processing of numerical magnitudes is automatic. In general, these paradigms are designed so that if number processing occurs, it has to be *involuntary,* in the sense of being irrelevant or even conflicting to the given task. In all four paradigms, the numerical distance effect is in focus: if numerical distance exerts its effect on the behaviour, or on the brain responses, than the meaning of numbers has been processed.

(1) Paradigm when the task does not require explicitly the processing of numerical meanings;

(2) Subliminal priming: when the presentation time of the stimuli is so short that participants cannot recognize (consciously);

(3) Adaptation: amount of behavioural response (e.g. looking time) or brain response (e.g. BOLD, ERPs or EEG synchronization) to 'familiar' decreases, brain response to 'new' increases;

(4) Numerical Stroop paradigm

(1) Dehaene and Akhaevin (1995) showed pairs of numbers written in different notations to their participants. The task was to match the two stimuli, either semantically or physically. For example, the stimuli pair '2 – TWO' had to be answered as "same" in the numerical matching task, and had to be answered as "different" in the physical matching task. Independent of the notation and independent of the task, the numerical distance effect was significant: participants responded slower when the numerical distance between than two numbers was small (e.g. '2 – THREE') than when it was large (e.g. '2 – EIGHT'). The meaning of these symbols and their relative positions on

the mental number line has been automatically accessed, even when a simple perceptual match would have been sufficient to solve the task. With a similar paradigm, involuntary access to numerical representations was demonstrated in 6-year-old children as well (Duncan and McFarland, 1980). Children's task was to tell whether the two numbers on the screen were physically the same or not (e.g. 4 4 – same; 2 4 – not the same). Even only a simple physical match would have been sufficient to answer this question, children showed the numerical distance effect: it was easier to tell that two number were not the same when they were further apart from each other.

(2) The *subliminal priming paradigm* provides an other way to demonstrate automatic access to numerical representations**.** In short, there is a 'prime' stimulus presented to participants first, which is followed by a 'target' stimulus requiring response. The prime stimulus should remain consciously unrecognizable for the participants in order to avoid any confounds possibly caused by the conscious strategies used by the participants. Using such a paradigm Naccache and Dehaene (2001a,b) showed that the meaning of a written number words was extracted and exerted significant effects on behaviour, ERP and BOLD signal even when participants did not perceive these words at a conscious level. In this experiment, participants were shown number words, but with such a sort presentation time (43 ms) which made conscious recognition of these words impossible. Following this priming stimulus and the visual mask afterwards, an Arabic digit appeared for 200 ms and participants were asked to decide whether it is smaller or larger than 5. They found that RT was faster when the prime and the target fell on the same side of 5, in comparison to when the prime and target fell on the opposing sides of 5. Furthermore, the effect of numerical distance was also significant: in congruent trials (when prime and target fell on the same side) the smaller was the numerical difference between the prime and the target, the shorter was the RT (see also Koechlin et al., 1999). With EEG and fMRI measurements Naccache and Dehaene (2001a) showed that the prime stimuli itself initiated a subliminal response in the motor cortex.

(3) Adaptation is the phenomenon when the amount of response, either behavioural or neural, decreases when a certain type of stimulus is repeated (also called 'repetition suppression'; reviews by Grill-Spector et al., 2006; Krekelberg et al., 2006).

Although the neural mechanism behind adaptation is highly ambiguous, adaptation is now a widely used technique in several fields of cognition, with the use of several types of measurements. Behavioural responses, e.g. the fastening and sharpening of reactions (e.g. Naccache and Dehaene, 2001a,b) or the reduction of looking time (e.g. Xu and Spelke, 2000) along neural responses, like the decrease in BOLD response or the decrease in ERP amplitude are claimed to indicate stimulus- or domain-specific reactions of the cognitive apparatus. Adaptation has been showed at the cell level for perceptual stimulation, at the level of circumscribed perceptual brain areas to somewhat more abstract perceptual stimulation, and again at the level of brain areas to categorical – conceptual categories, like face, word meaning or numerosity (for review, see Grill-Spector et al., 2006; Piazza et al., 2007).

If numbers were represented in a domain-specific manner, than we would expect the responses of participants to *adapt* to numerical information. Behaviourally, with a looking time paradigm, Xu and Spelke (2000) showed that infants indeed extract abstract numerical information from repeated dot displays which varied in perceptual properties but was kept constant in numerosity (This paradigm will be introduced in detail in chapter IV). Adaptation of the BOLD signal (measured by fMRI) to numerical stimuli was tested by Piazza et al. (2004, 2007) and Cantlon et al. (2006). Piazza et al. (2004) employed a paradigm in which reaction to change in number and in shape was contrasted. They repeatedly presented a specific number to participants either in symbolic (Arabic digit) or in non-symbolic (dots and triangles) notation. After this habituation stream, the represented magnitude changed unexpectedly (see Figure I/4).

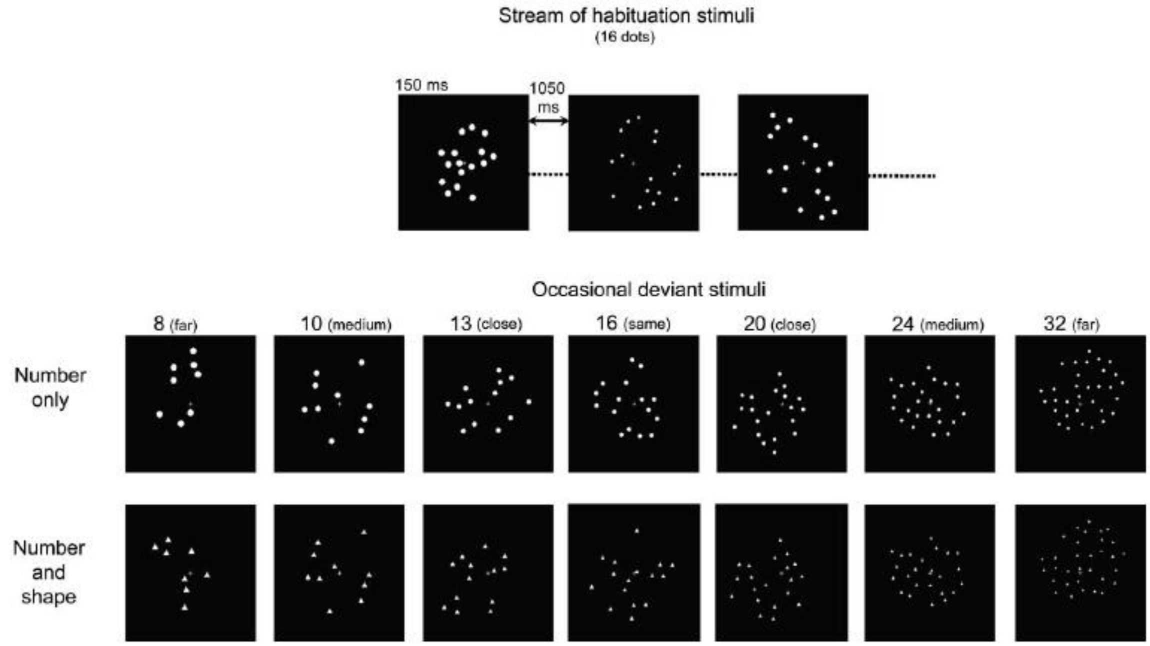


Figure I/4: Illustration of stimuli from Piazza et al (2004). Participants passively viewed a stream of displays like pictured in the figure. Occasionally, the number of items or the shape of the items changed; BOLD responses to number and to shape change were measured.

They found a significant rebound in the activity of the IPS areas to numerical change, which response was proportional to the change in numerosity (distance effect).

From the critical point of view, however, there are some unresolved design issues with these fMRI adaptation paradigms. First, in the Piazza et al. (2004) study although the surface and density of the dot displays were attempted to vary independently of number, the circumference indeed still correlated with number change. The authors also note that for example the numerosity of the display was approximated always a bit more when it consisted of triangles than when it consisted of dots. What they miss to report is that, in fact, when a triangle and a circle have the same surface size, the triangle has a larger circumference. This underlines again, as it has been already discussed in section 2.2.1, the importance of perceptual control in experiments with non-symbolic numbers. In their recent experiment (Piazza et al., 2007) it is noted in the methods section of the paper with small caps that: “To avoid decision and response confounds, participants were simply instructed to pay attention to the quantity conveyed by the stimuli. They were informed that they would be shown quantities in different formats and that their approximate values would be ~20 and ~50". Well, as soon as bringing the 'purpose' of the study into the attention of participants, conscious strategies and simple, conscious *change detection* cannot be disentangled from adaptation any more. In sum, the fMRI adaptation paradigms contain either perceptual confounds (e.g. circumference; Piazza et al., 2004; Cantlon et al., 2006) or confounds from task instruction (Piazza, 2007), both probably leading to a detection of change instead of adaptation.

(4) The *size congruity effect* obtained in numerical Stroop paradigms is the fourth evidence of the automaticity of number processing. For instance, Henik and Tzelgov (1982) and Tzelgov et al. (1992) found that numerical meaning interfered with physical size, during physical size comparison task (see section 2.3.). The intrusion of conflicting numerical meaning despite participants' obvious interest and effort in suppressing the evaluation of numerical magnitudes clearly demonstrates that numerical meaning is processed in an unintentional and automatic fashion.

2.6.1. The development of automatic number processing

It has been also shown that symbolic distance between numerosities affects 6-year-old children's performance even if numerical magnitude is not relevant for the given task, suggesting automatic access to refined numerical representation in young children. For instance, during a simple decision whether two stimuli on the computer screen are perceptually identical or not, responses slow down when numerical magnitudes are closer to each other in comparison to when they are further apart (Duncan and McFarland, 1980). However, using the numerical Stroop paradigm, when the physical sizes of the stimuli has to be compared, symbolic distance effect is not consistently found either in adults (Tzelgov et al., 1992) or in children (Rubinstein et al., 2002) in behavioural experiments. In a previous study (Szűcs and Soltész et al., 2007) using not only behavioural, but EEG measurements as well, we have already shown that grade 3 and 5 children process numerical meaning involuntary, showing NDE in event related potentials (ERPs) at around 210-230 ms after stimulus presentation.

➡ The developmental pattern of automatic number processing will be explored in younger children (grade 1 to grade 3) in Chapter VII.

2.7. Speed of access

Because numerical symbols (number word, Arabic digits) convey no perceptual information regarding the relative magnitudes and numerical distances, this information has to be abstracted. So forth the appearance of the *numerical distance effect* doubtlessly indicates that the abstract meaning of numerical symbols has been processed. Event-related potential studies have shown that the numerical distance effect appears already around 180-200 ms after stimulus presentation during number comparison tasks, both in adults (Grune et al., 1993, Dehaene, 1996) and in 5-year-old children (Temple and Posner, 1998) and in DD adolescents (Soltész et al., 2007). When numerical processing is involuntary, namely in the physical version of the numerical Stroop paradigm, numerical distance effect could be measured at around 200 ms in adults (Schwarz and Heinze, 1998) and in grade 3-5 children (Szűcs, Soltész et al., 2007).

The speed of access to numerical representations during involuntary number processing will be explored in younger children (grade 1 to grade 3) in Chapter V and VII.

II. Methods: electroencephalography, event-related potentials and principal components

Abstract

This chapter introduces electroencephalography (EEG) and the event-related potential (ERP) technique. Advantages and shortcomings of EEG and ERP are discussed. I will describe the specific ERP processing steps I developed and used in the ERP experiments of the thesis. Lastly, principal component analysis (PCA) will be introduced, as one possible method for the extraction of ERP components.

1. Background: Electroencephalography

Electroencephalogram (EEG) is the electrical activity generated by the brain, measured by electrodes placed on the scalp (Berger, 1929). Fortunately, there are quite a few excellent textbooks available (e.g.: Rugg and Coles, 1996; Handy, 2004; Luck, 2005; Andreassi, 2006; Nunez and Srinivasan, 2006; Cacioppo, 2007; Gazzaniga, 2008); so forth I would not like devote too much space in the thesis for the introduction of basic terms and concepts. In short, electrodes placed on the scalp detect the summation of several neurons' more or less synchronized post-synaptic activity. Post-synaptic potentials are the voltages that arise when ion channels open and close at the cell membrane after neurotransmitters bind to receptors on the cell. Post-synaptic potentials should not be confused with action potentials, although the two are closely related. The summation of excitatory or inhibitory post-synaptic potentials can result *in action potential*. However, the action potential is an all-or-none transient alteration and the saltatory manner of its conduction along the axon makes it difficult to measure further away of the given cell. Meanwhile, *post-synaptic potentials* are graded potentials, and form tiny dipoles along the cell. When a large conjunction of cells, aligned optimally in a parallel fashion, is activated approximately at the same time, these tiny dipoles summate and can be detected at the scalp. In sum, Electroencephalography (EEG) is a non-invasive method for measuring brain activity (summed post-synaptic electric potentials of several neurons activated in synchrony at the same time and oriented in a way that allows the electrical flow to reach the head surface).

1.1. Event-related potentials

One conventional method for the signal extraction from EEG data is the averaging of the data along time windows corresponding to experimental events. For the averaging procedure, first time windows are extracted from the EEG. These time windows are called epochs and are time-locked to certain events (like the appearance of a stimuli or response) of the experiment. The epoching procedure yields EEG segments, or single-trial waveforms, which are related to an event. These single-trial waveforms, *event-related potentials*, are then averaged across time points in order to gain averaged[4] event-related potential (ERP) waveforms. These waveforms consist of a sequence of positive and negative voltage deflections, which are called peaks, waves, or components. This sequence is thought to reflect the flow of information in the brain[5].

This averaging technique is a signal processing technique ideally enhances signal and eliminates noise. This technique presumes that the signal is constant throughout the experiment, while noise is random; so forth, averaging it would zero the noise out while keeps the constant signal. In practice, unfortunately, noise is never completely cancelled. Also, to gain a reasonable signal-to-noise ratio (SNR), tedious work is needed: SNR becomes better in proportion to the square root of the number of samples – for example to double the SNR, the number of samples needs to be quadrupled.

---

[4] The term *averaged* is skipped in most of the studies and I will skip it too. When talking about ERP, the averaged ERP should be meant. The non-averaged ERPs are mostly referred to as single-trial ERPs.

[5] Averaging event-related potentials is not the only way of signal extraction. For example, averaging keeps *evoked* responses while misses *induced* activity, which is the continuous background activity of the brain, not necessarily locked to the stimulus events; oscillatory activity is also neglected by the ERP method. For example, time-frequency analysis can reveal oscillatory activity, both *evoked* and *induced.* The more elaborated discussion of these methods are unfortunately beyond the scope of this thesis.

1.2. ERP components

Components of an ERP waveform can be described by polarity (positive or negative) and timing, by their scalp topography and by their supposed generator in the brain. However, as ERP components are supposed to reflect cognitive processes, the psycho-physiologically meaningful and functional properties of these components are relevant as well (Donchin et al., 1978, Picton et al., 2000). For example, all the parameters of a functionally identical component are subjects to variability under certain circumstances. For instance, the scalp topography of an auditory elicited N1 component depends on the pitch of the eliciting stimulus because of the tonotopic arrangement of the auditory cortex. Polarity may also vary: a component elicited by visual stimuli in the primary visual cortex swaps polarity depending on the visual field, because of the folding pattern of the cortex in this area. For a last example, the widely researched P3 component's latency ranges from 300 to 800 ms, depending on task requirements. And despite the variability of these superficial features, like topography, latency or polarity, these components reflect the same perceptual or cognitive processes in the brain.

It is not trivial how one can extract the functionally meaningful components of an ERP waveform. Actually, an ERP waveform can be decomposed into infinitely many possible combinations of underlying latent components. The observed ERP waveform is a summation of several processes, overlapping in space and in time; a voltage peak in fact may not reflect a 'peak' of a cognitive component or function. Overlapping *latent components* can add up and shape waveforms which may not be reminiscent of the latent components themselves (for an illustration, see Figure II/1).

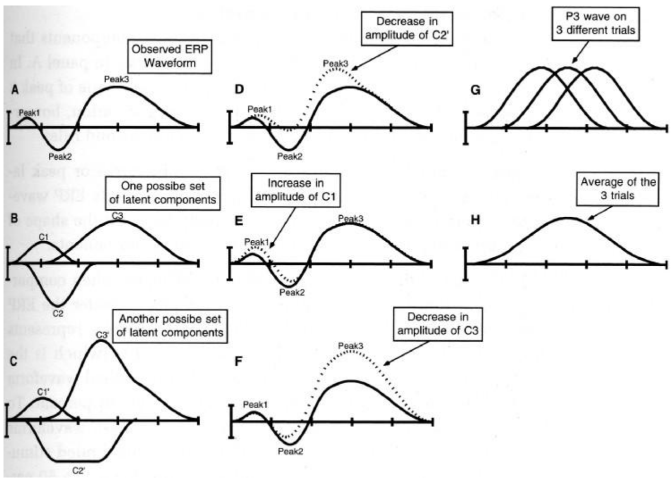


Figure II/1: An illustration of latent components, taken from Luck (2005). A: an observed component. B and C: possible solutions of latent components, which add up to the observed waveform. D, E and F: shows the observable results if the latent components in panel B are changed. G and H: an example of how could several latent components appear as one component in the measurements.

Latent components can be derived in several ways. For example, the Fourier transform decomposes the EEG/ERP signal into a set of sine and cosine waves (theoretically, though, any wave could be decomposed into square waves as well). Principal component analysis (PCA; see later) or the independent component analysis (ICA) is both correlational (linear and non-linear) methods for component extraction.

1.3. Advantages of the EEG/ERP methods

The biggest advantage of EEG in comparison to mere behavioural studies is that it does not require either conscious processes or overt responses from the subject. Clearly, by mere behavioural measures, we were not able to monitor ignored or unconscious processes; and both are highly relevant for example in the research of attentional processes. Furthermore, while the response or the reaction time is an end result of several preceding processes, EEG/ERP is a continuous measure and is able to capture several different processing stages leading to the end result. For example, the elongation of reaction time (RT) in an experimental condition in comparison to an other experimental condition could reflect that more time is needed for the perceptual processing of the stimuli, but it could also reflect that the decision was harder, or the response organization was more difficult. From RT alone, we would not be able to separate among these processes.

The high temporal resolution of the EEG/ERP method (in the range of milliseconds) ensures its superiority above imaging techniques (fMRI, PET) which rely on the pace of the blood flow in the brain, which takes several seconds.

1.4. Disadvantages of the EEG/ERP methods

One disadvantage of EEG/ERP is that it requires a large amount of data collected under strictly controlled laboratory conditions in order to disentangle signals from noise in the range of microvolts. Usually, more hundreds of trials are needed, what makes experiments sometimes tiring, boring and unnatural. Furthermore, if we consider an experiment lasting for 60 minutes when a data sample (voltage value) is taken in every millisecond at 32 electrodes on the head, we end up with 115,200,000 data points per each subject. This is an enormous set of information. Although the storage of such an amount of information is not a big deal any more, thanks for the previously unseen capacity of modern computers, the extraction of meaningful information is indeed still a challenge (the ERP technique is only one of the several methods for data reduction).

An other drawback of the EEG/ERP methods is that it does not provide spatial solution for the source(s) of the signal. EEG/ERP has a great temporal resolution, but on the flip side, it has virtually no spatial resolution. Voltage patterns measured on the scalp are the result of the summation of voltages of probably several dipoles, which activities overlap both in space and in time. Via volume conduction, electricity from these dipoles spread in all directions in the brain. And due to the high resistance of the skull, when the current flow meats the skull, it gets largely distorted – the surface distribution of voltages is smeared. Out of these reasons, to find the source(s) of a given surface pattern is an ill-defined, inverse mathematical problem (e.g. both 4+1 and 10-3-2 equals 5. If only the end result is known, there are an infinite number of solutions to get to this end result).

2. Methods: ERP

2.1. Data collection

EEG data was collected using the EGI's system. A 65-channel Geodesic Sensor Net (GSN, Figure II/2) is put on participants' head. The electrodes are embedded in sponges, which have to be soaked in salty water prior to application, in order to establish an optimal contact between the head and the sensors. Impedance has to be below 50 kΩ for the GSN in order to acquire good quality data. The analogue signal is amplified by an EGI Amp 200 amplifier. The analogue data is digitalized at 500 Hz and is recorded by the NetStation software (EGI). Behavioural data (response time and the code of the pressed button) is co-registered with EEG. The Presentation software (Neurobehavioral Systems) is used for stimuli presentation.

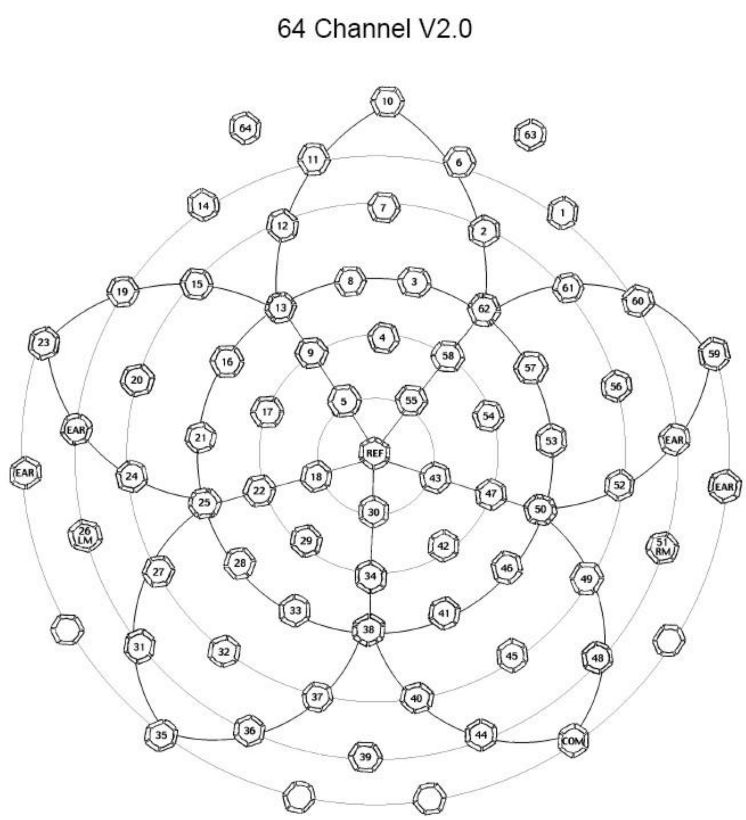


Figure II/2: Geodesic Sensor Net electrode arrangement. Nose is up, ear tokens are noted on the left and on the right side of the net. 'Ref' electrode in the middle (the reference) is above the Cz point according to the 10-20 system. Otherwise, the arrangement of the electrodes is not exactly following an extension of the 10-20 system; the net is rather designed to incorporate equally spaced sensors.

The binary data file is loaded into MatLab (MathWorks). From this point on, MatLab is used for all sorts of data processing and analyses[6]. Statistical analyses of composite values (see later) and post-hoc comparisons were done with Statistica (StatSoft Inc.).

2.2. Pre-processing

During pre-processing, the data is *epoched*, with the epochs being locked to the appearance of the stimuli (for the calculations of the lateralized readiness potential, see Chapter N, the epochs are locked to the response). Then, the data is *re-referenced* from Cz to average reference. There are of course several other options one could choose from for the site of reference (e.g. earlobes, nose, chin, mastoids, Cz, etc.); following the convention of our labs[7], I use average reference. Then, the *baseline* is removed from the epochs. The -100 ms to 0 ms range serves for baseline (for the response-locked epochs as well).

2.3. Data filtering and artefact rejection

EEG data is noisy. The values of the data are in the microvolt range (one millionth [10^{-6}] of a Volt), both external and internal noise interferes easily. These artefacts emanate from movement, blink, sweat of the skin, external electro-magnetic noise, etc. There are various methods for filtering or for rejecting these artefacts. The choice of methods depends on the actual researchers' decisions and on the quality of the data. There is no generally accepted and used 'list of operations', either a 'list of advises'. For example, the simplest way of filtering is the detection of extreme values. For instance, epochs containing values below or above ±100μV, are rejected. Or, certain patterns can be searched and eliminated from the data. Such bootstrapping methods are available in several commercial software, for instance for the recognition and for the elimination of

[6] The programme scripts for EEG processing were all written by me and by Szűcs Dénes. Data loading and EEG pre-processing scripts are the results of shared work. Filtering programmes and visualization for filtering, multi-way point-by-point ANOVAs and principal component analysis used in this thesis were all written by myself.

[7] MTA Pszichológiai Kutatóintézet - Fejlődés-pszichofiziológiai Csoport, Budapest; Centre for Neuroscience and Education, Cambridge

eye blinks. Or, one can scroll through all the data and pick noisy epochs by eye. Or, independent component analysis (ICA) can also be used for blink detection. ICA decomposes the data into independent sources of variance; from the independent components, one can choose which s/he regards as blink – then the data can be recomposed, but without this blink component. Filtering with ICA is an attractive but not yet widely used filtering method. I think it is largely subjective what one picks from the several components as blink. For this reason, I prefer to use filters which rather rely on algorithms and not on the subjective evaluation of the observer.

Data filtering is especially important in case of data from young children. Unavoidably, children move and blink more than adults. Also, they get tired and get bored earlier. Actually, with the use of relatively loose criterion, for example rejecting trials containing voltage values exceeding ±100μV, more than half of the children's data can be usually completely discarded. This leads to the loss of a significant amount of data, massively lowering the power of the study (and not to mention the waste of many restless hours spent with testing children…).

There are solutions which keep the contaminated epochs after 'cleaning' them. The before mentioned ICA method can be such a solution. Spectral filtering can be an other: certain frequencies, regarded as noise, can be subtracted from the data. In this procedure, the signal is decomposed into sinusoids. As mentioned before, any continuous signal can be decomposed into a set of sine (and cosine) waves with different wavelength and phases. Certain frequencies associated with for example muscle movement (~40Hz) or associated with tiredness (8-12 Hz), or with a slow drift of electrodes (~1 Hz), can simply be subtracted from the data. However, one cannot be completely sure that these frequencies contribute to 'noise' only. Actually, removal of frequency ranges does change the data remarkably, as we subtract a continuous and regular signal from the data. For example, taking out ~1 Hz can significantly lower the amplitude of a P3 ERP component in ERP measures. Out of this reasons, I prefer to reject contaminated epochs instead of the removal of a hypothetical component from the data.

2.3.1. Artefact rejection – in practice

The artefact rejection methods are partly based on the re-written routines which are implemented by EEGlab software (an open source code software running under MatLab; Delorme and Makeig, 2004; Delorme et al., 2007)[8]. Modified scripts of EEGlab's filtering package and some custom written code was used. The included filtering methods are the following.

(1) *Tresholding.* The simplest (and many times in fact the most effective) way of cleaning the data is to reject epochs which contain any data points below or above the specified threshold (e.g. ±100μV). Such high (or low) voltage values are mostly due to blinks and movements of the head or of the electrode(s).

(2) *Spectral rejection.* With this method, certain pre-specified frequencies considered as noise are checked: if the power of a certain frequency range in an epoch exceeds let's say 3 standard deviations from the average, the epoch is rejected. Please note the difference between spectral *filtering* (mentioned above) and spectral *rejection.* In spectral rejection, the contaminated epoch is rejected; in filtering, the to-be-filtered aspect of the signal is removed from the epochs, while the epochs are retained.

(3) *Improbable data.* The joint probabilities of all data points are calculated: how probable they are, relative to the surrounding data points. Simply speaking, a sudden big change in voltage from one data point to an other is considered to be improbable.

(4) *Abnormal distribution – kurtosis.* It is also a probability measure and reflects the 'peakedness' of the probability distribution of the data.

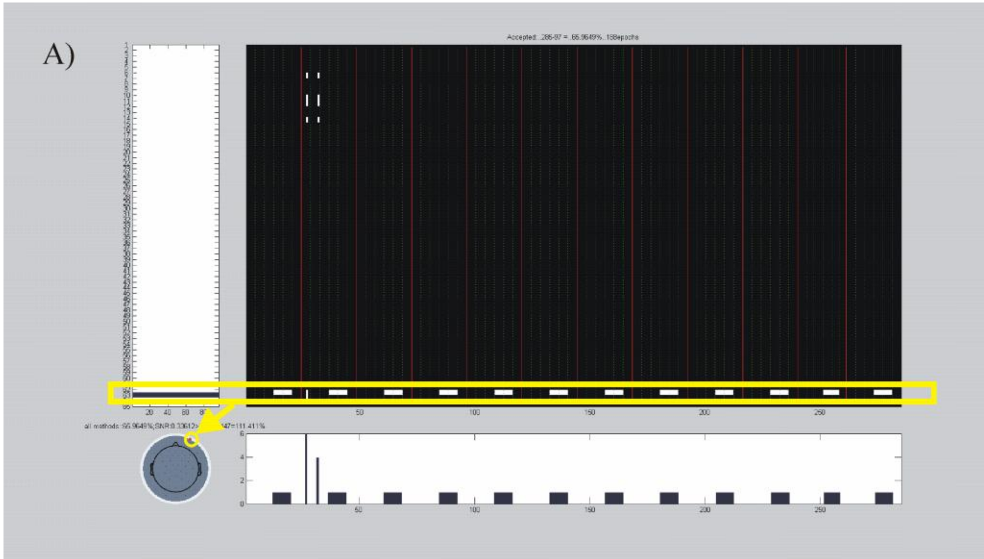
(5) *Linear trend.* Filters slow linear drifts in the data (transient recording-induced current drifts at low frequencies).

In my experience, basic *tresholding* and *spectral rejection* of alpha power (8-12 Hz) in children proved to be very efficient. Probability measures appeared to be not as efficient in terms of calculation time and the quality of the filtered data. Also, it seemed

[8] I have also reported some bugs in EEGlab's filtering package; these notes and suggested modifications have recently been considered and replied by Delorme and were put online (https://sccn.ucsd.edu/eeglab/bugzilla/show_bug.cgi?id=760)

to be difficult to find the optimal parameter settings for these filters, as there is not much precedent in the literature for these methods.

Electrode interpolation is a method to replace bad, noisy channels (electrodes) via spherical spline interpolation (a function in EEGlab) of data from the surrounding channels. If an electrode loose contact during an experiment or an experimental block, it causes the loss of several epochs during artefact rejection. When the noise comes from the lost contact of one or two electrodes, and not from other sources, it is indeed worth interpolating that channel. Figure II/3 and Figure II/4 illustrates two different instances of noise, one of which emanates from blinks while the other originates from a bad channel. I call blinks (or other artefacts related to movement or sudden, transient changes) '*vertical noise*' because it appears on the plot across several adjacent electrodes (and electrodes are the vertical dimension in the plot). And I call noisy electrode(s) as '*horizontal noise*' because this noise appears along one or two electrodes across several time points in the plot (time points are the horizontal dimension of the plot). In my experience, this visualization method is proved to be extremely useful, especially in children. The pattern, or distribution of noise tells a lot about the nature of the noise and may help decisions like whether it is better to drop the whole data set of a given child, or it is only one or two electrodes which have to be interpolated.



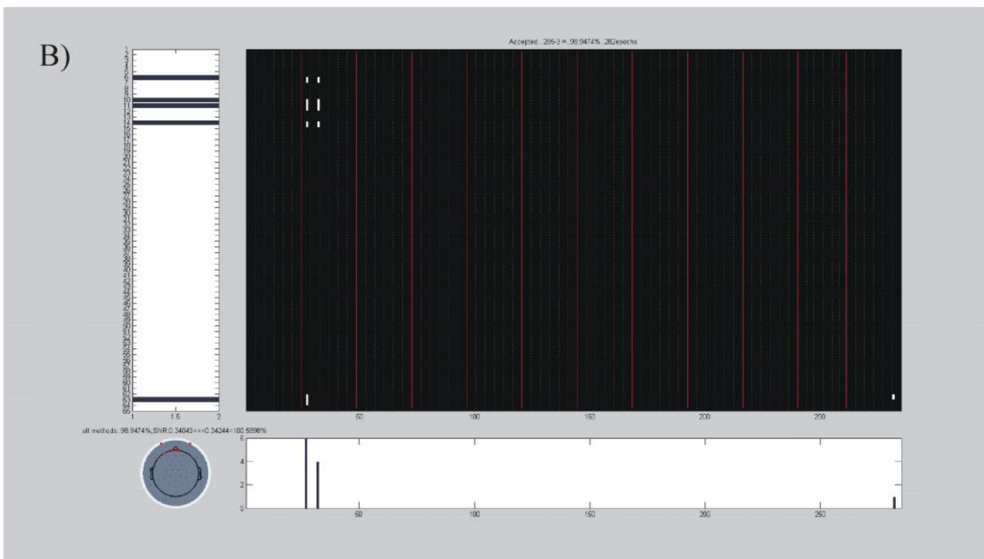
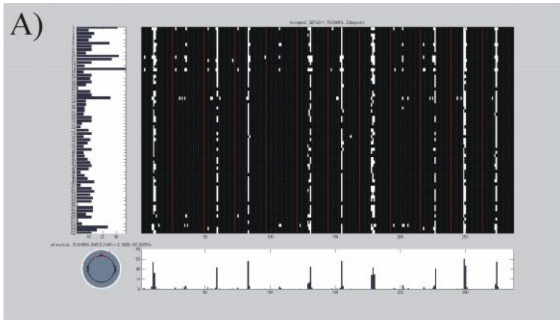
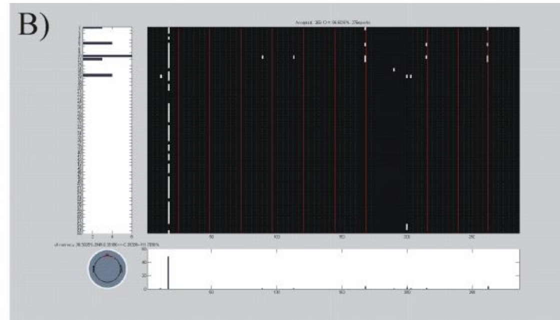


Figure II/3 and II/4 (next page): Composite figure representing noise (according to tresholding at ±150µV) across epochs in time (x axis) and across electrodes (y axis). Red vertical lines denote boundaries of experimental conditions. Dotted green vertical lines denote boundaries of experimental blocks. White marks denote epochs being rejected by the current rejection method. Insert figure on the left: represents the number of rejected epochs on each electrode. Insert figure on the bottom: represents the number of rejected electrodes at each epoch. Headplot in the left bottom corner: shows electrodes with noise larger than 3 standard deviations.
A: 'Horizontal noise' in one subject's data. Electrode 63 lost contact during the 4th and 5th experimental blocks.
B: The noise has been successfully eliminated by interpolating electrode 63.



II/4A: 'Vertical noise', or blinks in the present case, in one subject's relatively noisy data.



II/4B: Blink in an other subject's visibly much less noisy data.

If noisy electrodes are interpolated prior to filtering, the ratio of accepted epochs so forth the signal-to-noise ratio also will be much more optimal. Figure II/5 illustrates the distribution of noisy epochs in one child’s data, before and after electrode interpolation. Grand average ERPs (across all epochs in all experimental blocks) of the interpolated electrode are also shown; further, an other example from the same child demonstrates how the signal-to-noise ratio (SNR) is enhanced after rejecting noisy trials. SNR is calculated in a point-by-point manner, dividing the mean of each time point’s voltage across all epochs by the standard deviation of each time point’s voltage across all epochs.

To sum up, *tresholding* and *spectral rejection* of high alpha power (above 3 standard deviations) was used during artefact rejection, after the interpolation of noisy electrodes. An illustration for spectral rejection will be shown in the next section.

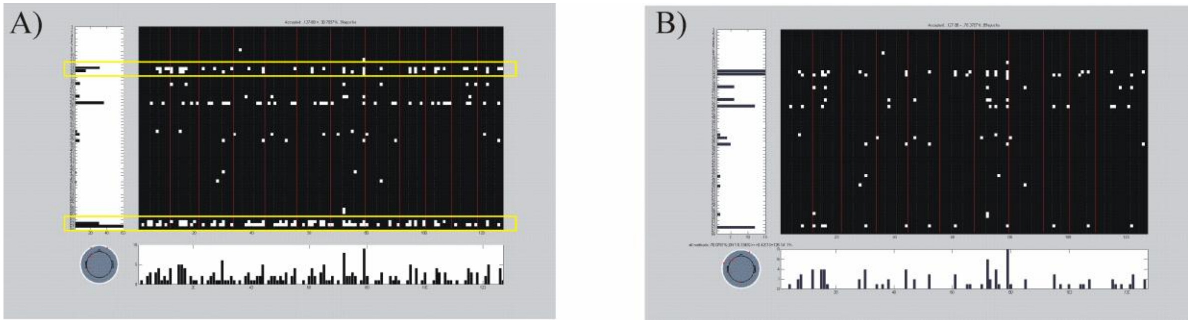
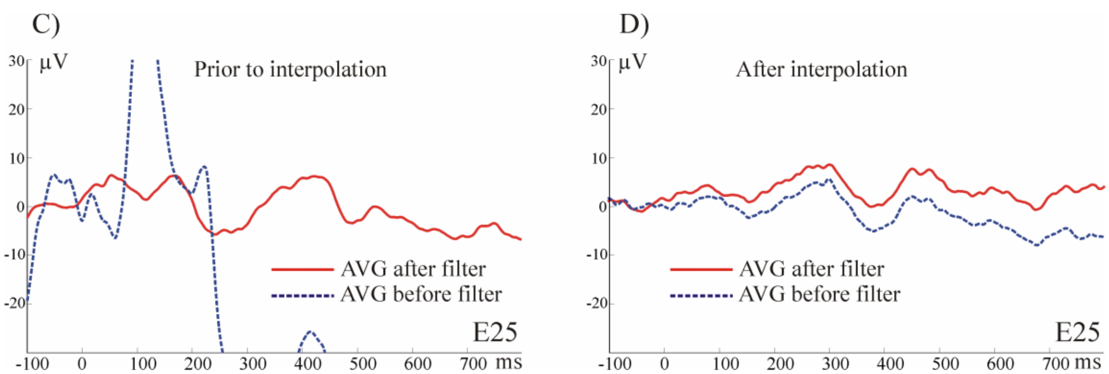
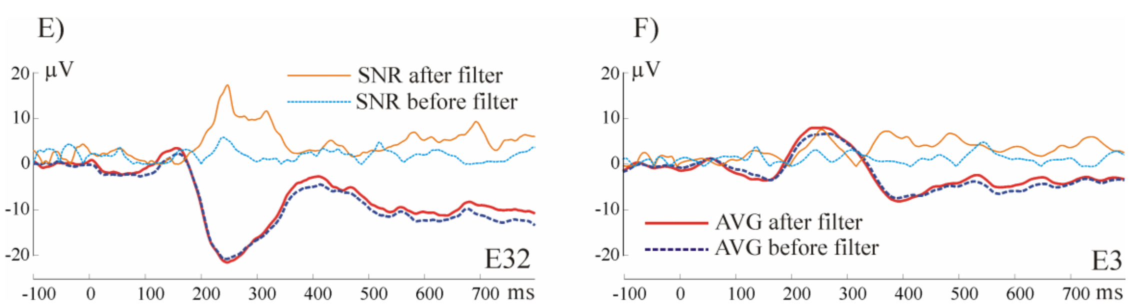


Figure II/5: One subjects's (children) data before (A) and after (B) interpolating four electrodes (in yellow boxes). The ratio of accepted trials increased from 30.7% (39 epochs) to 70% (89 epochs). Epoch rejection method: tresholding [±150μV].



The same children's data's grand average (AVG) of all epochs on one of the interpolated electrodes (E25) before (C) and after interpolation (D). Epoch rejection method: tresholding [±150μV].



Grand average (AVG) of all epochs and signal to noise ratio (SNR) before and after filtering the data, at two typical electrodes (same subject as above).

2.4. Data analysis: point-by-point analysis of variance (ANOVA)

The testing of experimental effects is algorithm-based, i.e. the process of definition and selection of experimental effect across time and electrodes is systematic and not subject to the biases of human eye observation, neither to the possible distortions of peak detection. There are 65 electrodes and 500 time points of ERPs; although an experienced observer can find relevant effects, picking electrodes and time windows by eye for statistical testing is rather subjective. Also, an effect does not need to be large (in size) to be significant – but still can be concealed from the observer. This is especially true when more than two conditions can be compared. *Peak detection*, when the maxima or minima of voltage deflections are searched and compared across experimental conditions, can also be deceptive or insufficient. As mentioned in section 1.2., peaks of voltage deflections do not necessarily reflect the 'peaking' of the underlying latent components. For these reasons, all data points (65 electrodes × 500 time point in each epoch, across all participants) are considered and subjected to statistical analyses.

There is though a problem with this point-by-point approach as well. So many tests of significance (500 independent tests in case of 500 time points) elevate the risk of false positives (type I error). There are different solutions for the correction of multiple testing, e.g. the Bonferroni method, where alpha is divided by the number of comparisons. This solution, however, yields an extremely strict condition for rejecting the null hypothesis (0.05/500=0.0001). In studies where point-by-point statistics were applied (e.g. Naccache and Dehaene, 2001a), usually a restriction regarding the number of consecutive time points and regarding the number of adjacent electrodes, which all have to show the same effect, proved to be reasonable. I will follow this convention.

2.4.1. Point-by-point ANOVA – in practice

For a point-by-point ANOVA, the ERP data is arranged into a multi-dimensional matrix of electrodes [65] × time points [500] × experimental conditions [C] × participants [N]. Then the F and p values for each possible effects are calculated (3 in case of a 2-way ANOVA (the two main effects and one interaction), and 7 in case of a 3-way ANOVA). The p values for either effect or interaction are plotted in the following way (see Figure II/6): black marks in an electrodes × time points matrix indicate if the p is below the pre-defined alpha level and if restrictions about the consecutive electrodes and time points are fulfilled. For the sake of concise representation, *spectral rejection* is also illustrated in this figure (Figure II/6, next page).

2.4.2. Statistical analyses and the reporting of composite values

A *Composite statistical analysis* is then performed in Statistica for the report of experimental effects. The time and electrode range of significant effects are read from this plots (Figure II/6) and extracted from the data matrix. The time intervals are averaged together so that instead of point-by-point values, one value per electrode is entered into the analysis. This data is then entered into an electrode × $effect_1$ × $effect_2$ … × $effect_n$ ANOVA[9].

Across electrodes, the polarities of voltage values, just as the direction of the experimental effects turn around. This is something what needs to be taken care of, because significant effects appearing with opposite directions at different electrodes annulate the main effect and yield and electrode × effect interaction. I tried two alternative ways to simplify these interactions.

(1) Electrodes are grouped into + and – sets and these sets are entered into two separate ANOVAs. The problem with this solution is that the direction of an effect can be swapped among electrodes even in the same polarity.

(2) An other possible solution is to simply examine the experimental effect at each electrode (when an electrode × effect interaction is significant); and then group the electrodes based on the direction of the experimental effects.

The results of these composite tests are reported and showed in ERP and topographic plots. Post-hoc comparisons for more than two conditions are represented in graphs.

[9] The examination of the ERP waveforms is of course inevitable if we are interested in whether the significant effects arise in the amplitude or in the latency of a certain component. To further test let's say the latency of a component, peak measurement is inevitable.

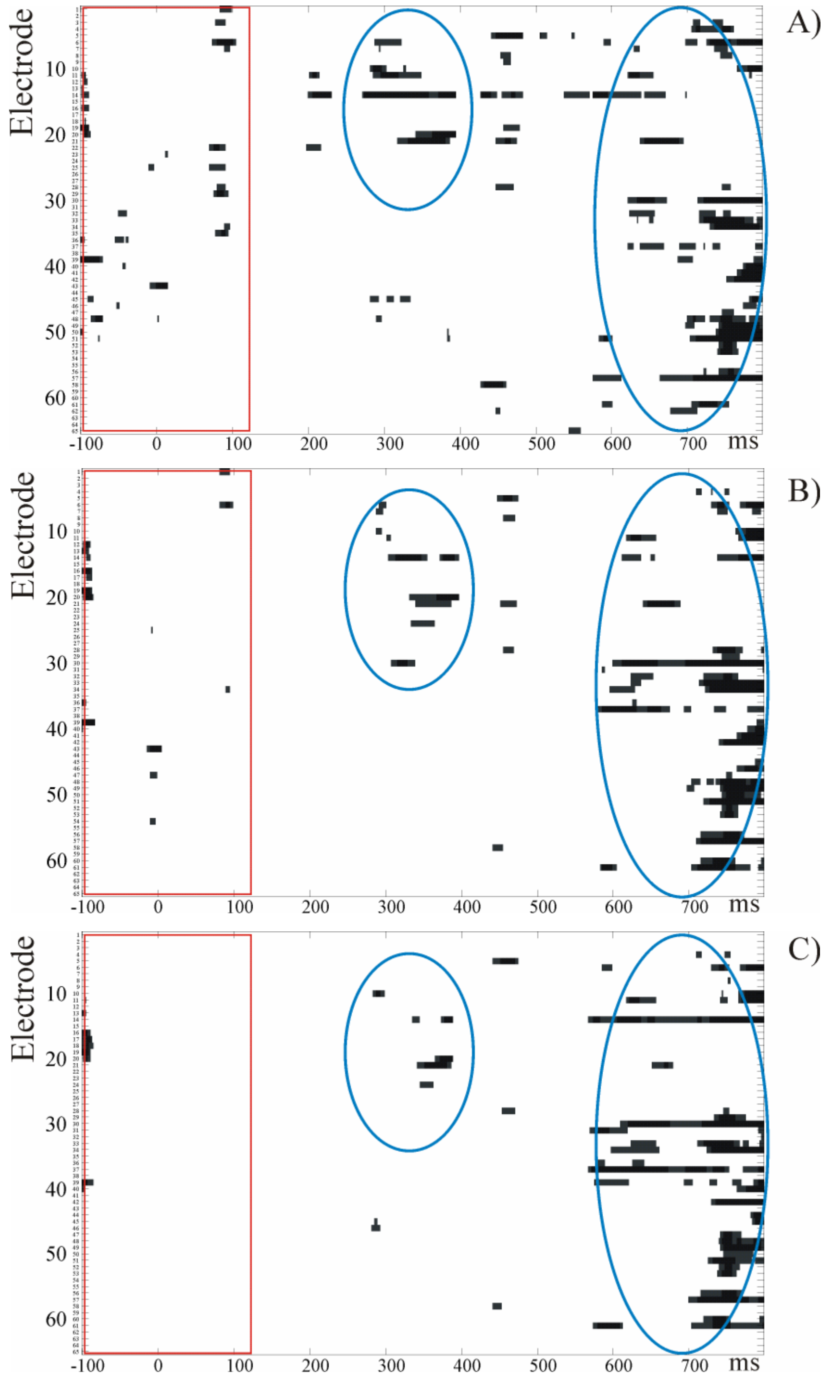


Figure II/6: Electrode × Time representation of significant effects with and without epoch rejection based on alpha (8-12 Hz) power, an example

A: Significant (at $p<0.025$) effects of an experimental manipulation, tested by a point-by-point within-subject ANOVA. 26 children's data were entered into the analysis. Red box: Most probably random effects (or stimuli locked oscillatory activity – noise in the present study). Blue ovals: experimental effects.

B: Epochs with a power of alpha exceeding 3 standard deviations are rejected. Number of epochs is reduced by 5%; SNR is enhanced by 203%; standard deviation is reduced by 38%.

C: Epochs with a power of alpha exceeding 2 standard deviations are rejected. Number of epochs is reduced by 11%; SNR is enhanced by 230%; standard deviation is reduced by 39%.

3. Methods: Principal component analysis

The principal component analysis (PCA) is a linear (based on correlational structure) method for data reduction. In EEG research it has been used to reveal latent components which compose the visible data.

PCA (Pearson, 1901, Hotelling, 1933; first application on ERP: John et al., 1964) belongs to the statistical data exploration methods family called factor analysis. The term 'component' might be confusing as, by convention in cognitive neuroscience, functional units of ERP waveforms are also labelled 'components'. In the statistical nomenclature, labels 'factors' and 'components' can be used interchangeably and both refer to the extracted vectors in the variance space. Throughout the thesis 'factor' and 'component' are also used interchangeably, with an attempt to keep ERP components and principal components as distinguishable as possible.

In PCA, it is assumed that there are a few (or, at least, fewer than in the original data) underlying components explaining most of the variance in the original data. To be more specific, in case of a temporal PCA there are hundreds of variables in the original data, namely all the time points of the original data (number of electrodes × number of sample points × number of subjects × number of conditions). It is obvious that we cannot see all the variability and interrelations among these time points (the number of relations among 200 time points is 19900!). Essentially, PCA reveals the complex relationships between the numerous dependent variables, based on their inter-correlations or co-variances. Co-varying time points are considered to belong together and form a basic waveform, or component. These components are extracted only if there is a variation across subjects, experimental conditions and electrodes, exactly what we are looking for in our data. Replacing all the time points in the original data PCA defines patterns of activity in time, the time points which are varying together will constitute to a basic waveform or component. All the separate basic waveforms or components have certain amount of activity in all the combinations of experimental conditions, subjects and electrodes. The differences in these activities among experimental conditions, subjects and electrodes can be analysed for each basic waveform. Similarly, a spatial PCA disentangles signals overlapping in space and defines latent components explaining common variances; so forth reduces the original dimensionality of the electrode space.

3.1. Component extraction and definition

PCA captures the variance in the data set by forming weighted linear combinations of the original variables. Computation of linear combinations by definition have the desirable mathematical property of orthogonality, providing ideal input into most of the statistical test requiring independent variables (like ANOVA). Components can be extracted by singular value decomposition (SVD; Golub and Reinsch, 1970; Harner, 1990, Harner and Riggio, 1989) from the mean corrected covariance or from the normalized correlation matrix of the original data. With the same measure across variables (microvolts across time points) the non-normalized co-variance matrix is preferred above the normalized correlation matrix (Boxtel, 1998, Kayser and Tenke, 2003). The SVD algorithm[10] decomposes the data matrix into three different matrices: into two unitary matrices and into one singular value matrix. One of the two unitary matrices contains the basic waveforms (the right eigenvectors, or factor loadings; for example the temporal vectors in a temporal PCA) and the other unitary matrix contains the information about the distribution of the basic waveforms (the left eigenvectors, or the spatial vectors in a temporal PCA). These two matrices are orthogonal matrices, meaning that the columns are orthogonal to (independent from) each other. The singular values are the square root of the variance for each temporal vector, the weight by which a given temporal factor constitutes to the original data. Having the original data set decomposed into these matrices, it can be reconstructed by a simple matrix multiplication of the three matrices. And that is the main purpose of PCA: the data can be reconstructed from only a few temporal factors, reducing the dimensionality of the original data. Speaking in spatial terms, the original data can be projected into a new space defined by the (temporal) vectors. For example, with 200 time points, the original

[10] Other methods, e.g. the NIPALS algorithm (Nonlinear Iterative Partial Least Squares) is also used for PCA computation. For example the software STATISTICA (StatSoft copyright) implements NIPALS decomposition during PCA. MatLab (MathWorks) provides SVD and eigenvalue decomposition (EIG) decomposition algorithms. EIG yields similar results to that of SVD, except for that EIG does not compute the left eigenvectors.

data had 200 dimensions in space. With, let’s say, 10 extracted temporal factors, the projected data has only 10 independent dimensions in space. It is obviously an effective way of data reduction, while retaining most of the information.

3.1.1. Selection of components

There are more methods for selecting the components which we want to retain for the re-composition of the original data. The best known criterion is the ’eigenvalue equals one’ rule, by which components having an eigenvalue larger than one are retained for further computations. In other words, these are the components which explain the variance of more than one variable (data point in our case). The eigenvalue equals one rule tends to overestimate the number of the relevant components (Boxtel, 1998). Another, most commonly used method in ERP PCA studies is the scree test (Cattell, 1966). In the scree test, loadings of the components are plotted as a function of their eigenvalues, in a descending order so that component with the largest eigenvalue represented on the left and the component with the smallest eigenvalue is represented on the right. Then, looking from right to left, the researcher defines an ’elbow’ point at wich eigenvalues start to increase. Although this method allows for subjectivity to a certain degree, data simulations showed that scree test results in more accurate number of retained components than eigenvalue equals one rule does (Boxtel, 1998). Other statistical methods were also developed for avoiding the subjectivity of the scree test and the overestimation of the ’eigenvalue equals one rule’.

3.1.2. Definition of component loading

Component loadings are the amount of contribution by a given component to the variables in the original data (a weighted linear combination of basic waveforms and their weights). It is the amount by which a component contributes to the voltage at each time point. In other words, it is the correlation of the component and the original variables and its value is between -1 and 1. When the covariance matrix is analyzed, loadings are covariances and need to be normalized to gain correlations.

3.1.3. Rotation of components

Varimax rotation criterion is used most often in order to find the simplest structure in the data by finding a solution in which the rotated factors retrain their orthogonality and have either large or small contributions (loadings) to the data, eliminating intermediate contributions. This rotation technique minimizes the overlap among the loadings of different components – exactly what we are aiming for. With other words, rotated factors will have the least overlap in time and will separate temporal factors from each other in an effective way (Pourtois et al., 2008, Chapman and McCrary, 1995).

Varimax rotation is the most commonly and successfully used rotation method. However, other, non-orthogonal or oblique rotation methods, for example the Promax rotation, are also exist and preferred by some (for example Dien et al., 2004, 2005). Varimax rotation can result in variance misallocation, or ‘leaking’, when similar and overlapping factors yielded by a PCA explain the same variance, inflating the type I error (Chapman and McCrary, 1995). This ‘leaking’ can be caused by the non-orthogonality inherent to the original dataset. Later statistical examination, e.g. an ANOVA can reveal these possible variance misallocations and one can account for these variance misallocations by retaining the component showing the experimental effect with high F value while discarding components with similar topographical pattern or time course but showing the same experimental effect with lower F values.

3.1.4. Definition of component score

Component scores represent the new data in the new factor space. After the components (vectors) are defined, the original data is projected back along these new vectors. Component scores are the estimations of the values on the given component, in each experimental condition, subject and electrode. If a covariance matrix was entered to PCA, the scores are in microvolt units and represent the difference from the grand mean (with a correlation matrix, the scores are unitless). Statistical comparisons of experimental conditions, or subjects, or electrodes are performed on the factor scores.

3.2. Implementation of PCA calculations and further analysis of principal components

The PCA processing steps are defined and implemented according to Tabachnik & Fidell (2007), Spencer et al. (2001) and Dien et al. (2005). Processing steps are executed by a custom-developed MatLab (R2008b) toolbox, using the native SVD algorithm for singular value decomposition.

Point-by-point ANOVA is performed also by a custom-developed MatLab program introduced in section II/t.

3.2.1. PCA calculations

First the data are transformed into a two dimensional matrix with variables as columns and observations as rows. Second, the data was mean centered, where column means are subtracted of the corresponding columns (eq. Z/1).

$$X0_{ij} = X_{ij} - \overline{X}_j$$

Where X0 is the mean centered data matrix; i denotes rows and j denotes columns. Then, X0 was decomposed by SVD, like (eq. Z/2):

$$X0 = U \times S \times V^T$$

Where U is the left orthonormal matrix of eigenvectors, V is the right orthonormal matrix of right eigenvectors (or component loading coefficients) and S is the diagonal matrix of component weights. '×' denotes matrix multiplication, different from simple multiplication (for the latter the • is usually used). '^{T}' denotes matrix transpose.

The covariance matrix is defined (eq. Z/3),

$$Xc = X0^T \times X0 = [U \times S \times V^T]^T \times [U \times S \times V^T] = V \times S^2 \times V^T$$

where matrix S^2 contains eigenvalues of components. As for covariance calculations, the normalization factor (multiplication by 1/(N-1), where N denotes the number of samples) is omitted as it does not influence the singular values decomposition. If not the mean corrected, but the mean corrected and normalized data matrix was used, the above calculation would yield the correlation matrix instead of the covariance matrix. In EEG research the covariance matrix is used, as the variables (the columns of the data matrix) are all measured by the same scale (μV) so forth do not require rescaling. Rescaling via normalization is more necessary where variables are measured by different scales (for example a data set consisting of reaction times and error percentages). Further, analysis of the covariance matrix leads to the extraction of principal components that correspond to the largest variance around the mean event-related potentials. Loadings and scores are interpreted as differences from the grand mean, in microvolts. While correlation matrix yields similar (though not identical) results, the measure is not retained and scores had to be rescaled by the standard deviation. Differences between PCAs carried out on the two matrices are not significant (Chapman & McCary, 1995), and results from covariance matrix PCA were found to be more reliable (Dien et al., 2005).

After decomposition, the component loadings are calculated via weighting[11] the right eigenvectors (V: component loading coefficients) by the corresponding weights in matrix S. Loading reflects the distribution and weight of each component by which it is represented in the original data. For plotting, the loadings were factored by the standard deviation in order to yield correlation values between -1 and 1 (when correlation matrix is used, the scale of the loadings is already between -1 and 1). By the nature of SVD calculations, the polarities of each component loading are arbitrary; both extremes (±1) represent high correlations with the original data. Signs of original data points (just as the original data itself) can be recomposed as the product of the three matrices (U, S and V).

---

[11] Although weighting is not necessary (e.g. Ferree et al., 2009). It does not change the shape or direction of loadings; it enhances relative differences among components with high loadings and components with lower loadings. Opinions differ whether weighted or non weighted loadings yield more valid results after factor rotation, as factor rotation is influenced by the relative weight of the input factors, by factors explaining larger variances are tend to be overemphasized (Dien et al., 2005).

As for the next processing step, component loadings are varimax rotated to produce a simpler component structure (eq. Z/4):

$$L' = R \times L$$

Where R is the rotation matrix and L is the loading matrix.

The final step is to calculate factor scores (eq. Z/5):

$$W = X0 \times B \text{ , where}$$
$$B = L' \times \text{inverse}(S^2)$$

Or simply without normalization:

$$W = X0 \times L$$

The obtained component scores (W) can be subjected to various statistical analyses.

3.2.2. Statistical analysis of principal components – a possible solution

After extracting the principal components which explain a considerable amount of the variance of the original dataset, the original dimensionality of the dataset is somewhat tamed (for example 10 spatial factors instead of 65 initial spatial factors, or electrodes), and redundant covariances of the original dataset are eliminated. However, the identification of factors (principal components) with *functional* ERP components demands further procedures. Further complicating the picture, if a sequential PCA of both temporal and spatial analyses was conducted, resulting in for example 8 spatial factors and in 10 temporal factors, these spatial and temporal factors have many possible combinations. These spatial and temporal factors are ordered according to the relative amount of variance they explain of the original data and are not coupled in any ways. Traditionally, the identification of factors with meaningful ERP components, or the coupling of temporal factors with their spatial counterparts, is analogous to the identification of relevant peaks and troughs of the ERP (see for example Spencer et al., 2001). In order to avoid possible distortions caused by the fallacies of visual inspection and a priori assumptions, variance analysis can follow the PCA. An ANOVA testing experimental effects can *functionally* identify principal components with ERP components (this is the so-called PCA-ANOVA approach; Chapman and McCrary, 1995). For example, Spencer and colleagues (2001) paired temporal factors with their supposed spatial counterparts on the basis of prior knowledge of temporal and topographical patterns of ERP components. The authors tested these a priori factor pairings only, and significant experimental effects confirmed their a priori identifications. However, the identification of factors is not always so unambiguous, and also, some factors do not resemble to any well known ERP component. Still, these latter factors may represent significant components of cognitive processing, but are concealed in an ERP so forth not known for the ERP literature.

One possible solution for the above problem is to run a point-by-point ANOVA on all the principal component scores and to identify all factors which are sensitive to experimental manipulations. For example, in the case of a spatial PCA, component scores are the time courses consisting of scores for each time point. Similar to the point-by-point ANOVA introduced in section II/t, experimental effects are tested at each data point across all electrodes (or spatial factors) and across all time points (or temporal factors). This ANOVA results in a matrix of F and corresponding p values, for each combination of temporal and spatial factors, for all main effects and interactions of the experimental design. For examining interactions among experimental manipulations and differences among experimental conditions in detail, ranges of data showing significant experimental effects are averaged and exported to Statistica software (StatSoft) for post-hoc analyses.

3.3. Possible applications and validity of the PCA method

Principal components can be extracted and analyzed across time points, where the spatial arrangement of the temporal factor scores are subjected to statistical tests later (temporal PCA, e.g. Pourtois et al., 2008). PCA can also be calculated across electrodes (spatial PCA, e.g. Donchin, 1966), when the temporal pattern of spatial factor scores are subjected to statistical analyses. Further, temporal and spatial PCAs can be combined in the following way: factor scores from one PCA are entered into the other PCA, resulting in only one score for each combination of spatial component, temporal component and experimental condition. Spatiotemporal PCA is first conducted across electrodes, than component scores of the spatial components are fed back into a PCA across time points (spatiotemporal PCA, e.g. Spencer et al., 2001). Spatial PCA proceeds temporal because scalp distribution of an ERP component is assumed to be a fix attribute, while the timing of a given component can vary across studies and experimental arrangements (Donchin et al., 1978). Reversely, temporal PCA can precede spatial PCA, arguing that spatial components tend to overlap due to volume conduction (Dien and Frishkoff, 2005). The procedure applied depends on the technical limitations, for example a spatial PCA should be handled cautiously when the number of spatial dimensions (electrodes) is below 64. Spatial PCA is often used for the purpose of separating than locating neuronal generators (e.g. Kayser et al., 2009), as spatial PCA is supposed to identify and separate generators which activities measured on the scalp is overlapping. The chosen procedures of course also depend on the research question to be answered. For example, spatiotemporal PCA separates components first in space, than in time. This method proved to be very efficient in paradigms where functionally different components were overlapping both in space and time (Spencer et al., 2001).

PCA is able to decompose ERPs into their constituent components and to disentangle otherwise overlapping and not discriminable waveforms. Referring back to the concept of components, as defined by Donchin et al. (1978), PCA provides an effective tool to reveal *functionally* different and *functionally* relevant components, which respond differently to experimental manipulations and might remain concealed after only a peak oriented ERP analysis. PCA of course is only one of the numerous EEG analysis methods and as such, does not guarantee unique and correct solutions. It's clear advantage over the most traditional methods, like ERP peak measurement, is that it is purely data-driven and is not subject to the researchers' subjectivity and to their a priory assumptions. Components are not defined by their morphological properties, like by the visually easily detectable high voltage hills and troughs. Rather, components are selected based upon their statistical properties. For example, when voltages are in a higher range, differences between experimental conditions are easier to spot by eye, as they are larger in physical size. However, when voltages are in a smaller range, differences between components are much harder to recognize. Similarly, one usually focuses on the maximas and minimas of ERP waveforms, while differences among experimental manipulations may occur outside of these maxima and minima.

There are other methods for EEG data decomposition, for example the independent component analysis (ICA; Makeig et al., 1999a,b) or the 'topographic components model' (Möcks 1988), among several others. Unfortunately, there is no space to elaborate all these methods in detail in the present thesis. However, we have to introduce PCA in a few words because ICA may provide a considerable alternative for PCA. Contrary to PCA, which is a procedure applying linear decomposition, ICA is a nonlinear decomposition method. A key assumption in linear decomposition methods is that the underlying latent factors of variance are orthogonal to each other. According to some researchers (Makeig et al., 1999a,b, Onton et al., 2006), the variance of brain activations measured on the scalp are cannot be assumed to be orthogonal to each other; so forth ICA offers a physiologically more valid decomposition than PCA.

In summary, PCA proved to be internally consistent, yielding coherent, meaningful and most importantly replicable results in several studies (e.g. Spencer et al., 2001, 1999, Dien et al., 2004, Kayser and Tenke, 2003, 2005).

3.4. PCA applications in the thesis

PCA was performed on two different datasets in the present thesis. A temporal PCA was conducted on the ERP dataset of the developmental dyscalculia study (chapter Y, Soltész and Szűcs, 2009) and a spatiotemporal PCA was performed on an adult dataset from a numerical Stroop paradigm (Chapter S). PCA provided information beyond and above the traditional ERP analyses in both cases. In the developmental dyscalculia study, PCA revealed some group differences not recognizable from ERPs. In the adult number Stroop study, PCA disentangled functionally different components in the 250-600ms time window, unseen for traditional ERP analyses.

III. Developmental dyscalculia

Abstract

First I review the most recent theories and investigations of developmental dyscalculia[12]. Having the modular theory as a departure point, pro and contra behavioural and brain imaging evidence are reviewed. Theories of domain-general deficits are also introduced. Second, a re-analysis of my previous ERP data (Soltész, 2004, Soltész, Szűcs et al., 2007) complemented with neuropsychological measurements and with a temporal principal component analysis is included. Third, the present findings are discussed in light of the theories presuming the impairments of more general cognitive abilities behind developmental dyscalculia and mathematical learning deficits.

1. Developmental dyscalculia: an overview

Developmental dyscalculia (DD) is a cognitive disorder "affecting the ability of an otherwise intelligent and healthy child to learn arithmetic" (with relatively high-prevalence, 3-6.5%) (Gross-Tsur et al., 1996). DD appears despite normal intelligence, proper schooling, adequate environment, normal socioeconomic status and motivation (DSM IV). Contrary to its prevalence similar to dyslexia and contrary to its doubtless relevance in education, employment and mental health of individuals, DD is by far not as documented as dyslexia. For example, the target word "dyslexia" yields 34 times more results than the target word "dyscalculia" in the PubMed (http://www.ncbi.nlm.nih.gov/pubmed/) database (6378 vs. 187 peer-reviewed papers). Not surprisingly, DD still lacks a generally accepted definition and the cognitive component processes contributing to arithmetic performance are still poorly defined.

[12] Some rather tangential issues to the present thesis, like the history of discalculia research and the debate over its behavioural classification methods are already reviewed in great detail elsewhere (see Márkus, 2005; Soltész, 2004).

DD, or *mathematical learning deficits* can take several different forms and can be generated or accentuated by various types of cognitive deficits. There are at least three different theories for the explanation of DD; all three postulates different diagnosis criteria, and offers qualitatively different training and remediation approaches for teachers and for special education.

1.1. The defective number module hypothesis

Currently the most widespread neuro-cognitive explanation of DD is that there is a dedicated, specialized *magnitude representation* in the horizontal intraparietal sulcus (IPS; Dehaene et al., 2004) and that this number-specific representation is impaired in DD (e.g. Ansari, 2008; Butterworth, 1999; Landerl et al., 2004).

The empirical basis on which this assumption builds mainly comes from the finding that DD children and adults are *slower* and more *error-prone* during magnitude comparison and arithmetic tasks (Landerl et al., 2004; Rubinsten and Henik, 2005; Landerl et al., 2009).

1.1.1. Evidence from structural and functional brain-imaging

Supporting the *defective number module hypothesis*, structural brain imaging also found abnormal parietal structure and activation patterns in subjects with arithmetic difficulties. Levin (1996) examined a subject with dyscalculia and found abnormal asymmetry in the parietal lobes during calculation compared to healthy subjects. While bilateral activation of the supramarginal regions were observed in healthy control subject, the left posterior parietal regions were activated dominantly in the dyscalculic patient. In an other structural imaging study, voxel-based morphometry indicated a decrease in grey matter in the left IPS in children with very low birth weight and also suffering from dyscalculia (Isaacs et al., 1999). Further, magnetic resonance spectroscopy showed hypometabolism in the left IPS in adult with developmental dyscalculia (Levy et al, 1999). Besides structural, some functional imaging studies also showed differences in brain activity between DD and normal subjects. During solving arithmetic equations, Fragile X syndrome patients showed bilateral prefrontal and only left angular gyrus activation compared to healthy adults where bilateral prefrontal and bilateral parietal activation was found. Furthermore, unlike in healthy subjects, the task difficulty and the level of activation were not correlated in DD subjects (Rivera et al. 2002). Subjects with Turner's syndrome failed to show any differences between left and right parietal activation in function of exact and approximate calculations compared to healthy adults (Molko et al, 2003). Kucian et al. (2006) fMRI study also found less activation in the left intraparietal sulcus during approximate calculation but not during exact calculations in DD subjects compared to healthy children (Kucian et al., 2006). Price et al. (2007) compared DD children with normally performing controls in a non-symbolic number comparison task and found that the IPS activity showed no distance effect in DD, while distance effect was significant in controls. Cohen Kadosh et al. (2007) using the numerical Stroop paradigm, found that DD adults showed interference but not facilitation, while normal adults showed both facilitation and interference. Further, the authors applied transcranial magnetic stimulation (TMS) to the parietal cortex in normal subjects and found that their performance in the number Stroop paradigm became similar to that of DD subjects. Askhenazi et al (2008) reported a case study of a patient who suffered from (acquired) acalculia following an infarct restricted to the left IPS. This patient showed deficits with basic number processing, counting, subitizing and number naming.

The above observations are in coherence with theories assuming that DD is coupled with abnormal brain structure and/or function (Shalev and Gross-Tsur, 2001), specifically of the parietal lobes (e.g. Butterworth, 1999; Dehaene, 1997), potentially arising as a consequence of genetic influences and heredity (DeFries et al., 1997; Alarcón et al., 1997; Shalev et al., 2001; Shalev 2004).

1.1.2. Methodological issues

The evidence listed above are convincing for the first sight. However, there are some problems with these arguments. For example, although DD children were found to be slower during a number Stroop task, they were slower in the non-numerical (size and luminance comparison) control tasks as well in the Rubinsten and henik (2005) study. Second, and most importantly, most of these studies did not find an abnormal *behavioural distance effect*, which would be the strongest evidence for an abrupt magnitude representation. In fact, in all the studies where distance effect was measured, it did not differ between the DD and control groups. Although Price et al. (2007) found that the distance effect was not significant in DD in the activity of the IPS, the behavioural distance effect was indeed significant and did not differ from the controls', making the assumed functional relationship between IPS and magnitude representation questionable. In the Askhenazi et al. (2008) study the acalculic patient also showed, although a somewhat larger than normal, behavioural distance effect during number comparison.

Third, some of the studies simply forgot to administer any control tasks besides mathematics and number manipulation (e.g. Price et al., 2007; Cohen Kadosh et al., 2007). The different from normal IPS activation accompanying mathematics and number manipulation tasks in DD cannot be considered as functionally and distinctively related to magnitude processing.

Fourth, executive control and attention may explain some of the results just as well as the deficient magnitude representation theory does. Using the numerical Stroop paradigm, it was found that facilitation effect is absent in DD (Cohen Kadosh et al., 2007; Askhenazi et al., 2008). However, the same pattern has been shown in children with attention deficit disorder (Kaufmann and Nuerk, 2006). Further, Stroop tasks involve the detection and resolution of both stimulus conflict and response conflict, so that executive functions and inhibition abilities are also prerequisites of good performance in such tasks (Szűcs and Soltész, 2007; 2008). In accordance with this latter observation, in a recent study where the authors have taken IQ into account, the IQ partially accounted for the differences found in the size congruity effects between DD and normally achieving controls (Landerl and Kölle, 2009).

Fourth, structural imaging studies which demonstrated grey and white matter reduction in DD (Isaacs et al., 2001; Rotzer et al., 2008) cannot prove any functional links between the IPS and number representation, let alone the fact that abnormalities in the brain tissue was not restricted to the area of IPS in these studies.

In sum, the common finding that DD subjects are slower and more error-prone than their control peers during numerical tasks can not in itself prove the *number module* deficit; being slow and erroneous can result from several other underlying reasons, and these signatures are not considered as markers of the magnitude representation. In many cases, only the number-specific results are reported, while there are visible differences e.g. in the general speed during certain tasks as well. For example, Rubinsten and Henik (2005) reported that DD children showed weaker or none size congruity effects during numerical Stroop tasks where greyness, height and size were co-varied with numerical meaning. They concluded that the automatic access to numerical meanings is deficient in DD children, while they disregarded the fact that DD children were much slower than controls even if greyness or physical sizes of stimuli had to be compared. Furthermore, a significant amount of behavioural, developmental and educational research has been neglected by the investigators of the *number module.* For example, attention, working memory and inhibitory control have been found as significant contributors to mathematic disabilities.

1.2. Deficit of domain-general abilities

There are domain-general assumptions regarding the aetiology of developmental dyscalculia as well. For example, some hypothesize the deficits of the working memory lay behind DD (Geary, 1993, 1994, 2005; Geary and Hoard, 2001; Hitch and McAuley, 1991; Passolunghi et al, 1999; Passolunghi and Siegel, 2001; Siegel and Ryan, 1989; Swanson, 1993; Barrouillet et al. 1997). Working memory, short-term memory and executive functioning have been found to predict mathematical achievement in normally developing 7-year-old children (Bull et al., 2008). *Working memory* refers to the mental ability of the temporary storage and manipulation of information. More specifically, the *central executive* of Baddeley and Hitch’s working memory model (Baddeley and Hitch, 1974) is supposed to be the most important working memory component in arithmetic performance, as it predicts mathematics achievement in normal, but low math achieving children (Bull et al., 1999, 2008; Passolunghi and Siegel, 2001) and in DD children as well (Roselli et al., 2006). Executive functions can be considered as an umbrella term for general cognitive functions supervising, controlling and managing other cognitive functions (e.g. attention, planning, cognitive flexibility, inhibition, error detection and correction etc). Effective inhibition of intruding wrong associations and the inhibition of prepotent responses are also found to be less efficient in children with DD: they commit more intrusion errors in working memory tasks, even with non-numerical verbal stimuli (Listening span task: Passolunghi et al., 1999; Passolunghi and Siegel, 2001, 2004).

A recent fMRI study (Rotzer et al., 2009) measured brain activity during a spatial working memory task (Corsi block tapping test) in 8-10 years-old DD children. DD children's behavioural performance was significantly weaker in this task than their control peers' performance, and they also showed weaker neural activation during this task in the right IPS, right insula and in the right inferior frontal lobe.

Attentional deficits were also found in DD children (Ashkenazi et al., 2009). Error monitoring is also found to be abnormal in children with DD: error-related ERP signatures of error monitoring were found to be different in children with DD from those of normal children during simple, and most importantly, *non-numerical* discrimination tasks (Burgio-Murphy et al., 2007). These results also suggest that inhibition and error monitoring deficits of MD children are general abilities and are not specific to numerical information.

Further, DD is often associated with the weaknesses of visuo-spatial abilities (Rourke et al., 1993; Rourke and Conway, 1997), and children with DD often show slight deficits in perceptuo-tactile capacities, which deficits predict 5-year-old children's numerical processing and calculation performance 6 months later (Fayol et al., 1998; Noël, 2005). Gerstmann syndrome, a neurological condition also involved dyscalculia together with spatial and finger representation problems, such as left-right confusion and finger agnosia (Gerstmann, 1940). The putative (either anatomical, or functional) link between finger representation and number processing has been also demonstrated by Rusconi et al. (2005), who applied TMS to the left angular gyrus and found that both magnitude processing and the finger scheme were abrupt in healthy participants (see Figure III/1).

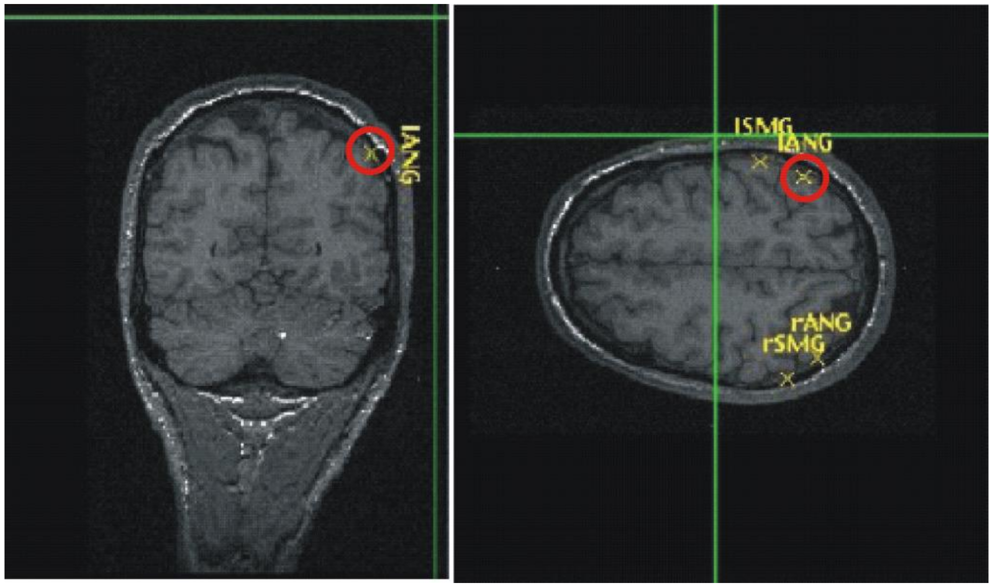


Figure III/1: taken from Rusconi et al. (2005). Front view and top view of a head showing the location to which TMS was applied, resulting in the abruption in both magnitude processing and finger representation.

Additionally, the co-morbidity of DD and dyslexia (17%) and attention deficit/hyperactivity disorder (ADHD) (26%) are relatively high, indicating the that a more widespread range of abilities might be involved in the aetiology of DD. These non-numerical abilities in relation with DD may well question the number-specific nature of DD, as it is assumed by the defective number module hypothesis (Butterworth et al., 1999; Landerl et al, 2004).

2. A combined event-related potential and neuropsychological investigation of developmental dyscalculia[13]

Abstract

Adolescents with developmental dyscalculia (DD) but with no other impairments were examined with neuropsychological tests and with event-related brain potentials (ERPs). A matched control group and an adult control group were tested as well. Behavioural and ERP markers of the magnitude representation were examined in a task where subjects decided whether visually presented Arabic digits were smaller or larger than 5. There was a normal behavioural numerical distance effect (better performance for digits closer to the reference number than for digits further away from it) in DD. This suggests that semantic magnitude relations depend on a phenomenologically (nearly) normal magnitude representation in DD, at least in the range of single-digit numbers. However, minor discrepancies between DD subjects and controls suggest that the perception of the magnitude of single digits may be slightly differed in DD. Early ERP distance effects were similar in DD and in control subjects. In contrast, between 400-440 ms there was a focused right-parietal ERP distance effect in controls, but not in DD. This suggests that early, more automatic processing of digits was similar in both groups, and between-group processing differences arose later, during more complex controlled processing. This view is supported by signs of decelerated executive functioning in developmental dyscalculia. Further, DD subjects did not differ from controls in general mental rotation and in body parts knowledge, but were markedly impaired in mental finger rotation, finger knowledge, and tactile performance.

---

[13] This Chapter, with some modifications, has been published: Soltész, Szűcs et al., 2007, Neurosci Lett.

This data also has been presented in a less advanced and less detailed form in my MA thesis (2004). Here I analyzed data on all the electrodes and across all time points. Further, instead of peak search, a point-by-point test was applied here. All distances, not only the two extreme, were also considered when testing the numerical distance effect. In behavioural data, an individual analysis was also conducted and slopes of the distance effect were also compared.

2.1. Rationale and background

In this experiment, DD subjects with no other diagnosed deficits apart from calculation disturbances were compared to control subjects. Since the DD participants were unimpaired in verbal processing and visuo-spatial skills, it is assumed that their calculation problems were most probably related to the dysfunction of the magnitude representation. Therefore I sought whether two signatures of the magnitude representation, the behavioural DE, and its electrophysiological correlate, the event-related brain potential (ERP) DE appears in our DD subjects. ERP distance effects are amplitude modulations of the parietal P3 ERP component (Grune et al., 1993), anIII/or amplitude modulations of parietally measured ERPs at around 200 ms after stimulus presentation (Dehaene, 1996; Szűcs and Csépe, 2005). The lack or an abnormal pattern of behavioural and/or ERP DEs in DD subjects would suggest that the impairment of the parietal magnitude representation underlies DD without verbal impairments (e.g. Butterworth, 1999; Molko et al., 2003). Bearing in mind alternative theories of DD I also examined finger knowledge, spatial skills, executive control and attentional skills of participants.

2.2. Methods

Subjects. Shalev et al. (2005) have demonstrated that DD persists throughout years and its presence is not correlated with socioeconomic status, gender, with the presence of other learning disabilities and with the extent and length of educational interventions (Shalev et al., 1998). In order to make the selection of subjects as valid as possible an independent special education institution was asked to select participants from their patient pool. The selection criteria were that subjects should have no other diagnosed deficit than dyscalculia, they have had participated and resisted therapy for at least two years, and had at least normal socio-economic status. Out of a pool of about 50 candidates 8 potential participants were identified. One boy ceased to cooperate during testing; this left 7 girls in the dyscalculia group (DG: mean age and standard deviation: 17±1.41 years). All subjects have had attended neurologists, and none of them were diagnosed with any neurological conditions. All subjects were highly functioning, wishing to pursue studies in humanities at university level. All subjects were attending mainstream public high schools but they had official statements of dyscalculia, i.e., they were not required to attend math classes. Each subject agreed to participate in three experimental sessions (one for EEG, two for behavioural testing). The IQ of DD subjects was determined previously during clinical assessment. Only the composite scores were available and, as experimental time was very limited, subjects could not be retested (IQ, VQ and PQ scores, standard deviations for DG and for CG consecutively: IQ: 98.17 (SD:13.53); 120.17 (3.54); VQ: 93.67 (8.46); 112 (5.37); PQ: 103.33 (24.51); 126.17 (4.17)). Therefore school performance was used in matching control subjects to DD subjects. The marks in DG were the following: Arts (literature, grammar, history, foreign language): 4.22±0.71. Sciences (biology, physics, chemistry, geography: 3.26±1.15). There was an adolescent control group matched in gender, age and school performance (CG, mean age: 16.43±1.39 years. Mean and standard deviation of marks: Arts: 4.18±0.77. Sciences: 3.71±1.08). One adult comparison group was selected (adults, mean age: 21±2.14; all university students).

Clinical investigations. Verbal, general, and basic mathematics skills of DG had been tested extensively by the special education institution. All DG pupils were slow and unsure in the naming and writing of numerals, especially with numbers above 100. All participants were slow and error-prone in both one and two-digit multiplications. 6 out of 7 DG subjects had difficulties in understanding and using place value. All subjects used fingers and other tools during counting. DG subjects recited math story problems accurately but they could not translate stories into mathematical equations. Their understanding of and memory for other than math texts and reading/writing achievement were normal. Interestingly, all DG subjects could recognize and apply simple rules with numbers below 20. (e. g. complementing the series “2, 5, 9, 14…” with the answer 20), though their solution time was very long. Participants could also understand and apply rules in non-mathematical contexts (e.g. continuing series with letters, or pictures of a scenario).

Neuropsychological tests. The neuropsychological test battery was selected according to former results and assumptions regarding DD. Attention was measured by the Toulouse-Pieron test. In this test children searched through, in a line-by-line manner, a set of printed squares with a small line attached to the squares. Out of a pool of 400 squares children identified squares with four predetermined kinds of line orientations. They had five minutes to find as many target squares as possible. Performance was calculated by dividing the number of the searched items by the number of errors. Short-term auditory memory and sequential processing was measured by forward and backward digit-span. Executive control was tested by the Trail Making Test (TMT), Part A and B (Arbuthnott & Frank, 2000). Finger representation were tested by a mental hand rotation task which consisted of 4×20 pictures of a hand rotated by 0º, 45º, 90º or 180º. Subjects decided as fast as possible whether pictures delineated left or right hands. In a finger naming task children named and showed their finger touched by the investigator with eyes closed and eyes opened (3×5 problems). Tactile-perceptual abilities were measured in a task where 10 everyday objects (e.g. key, pen) were identified separately with both hands with eyes closed. Body representation was examined by a task where children pointed to their body parts named by the experimenter (5 problems). Mental rotation was tested by Shepard's mental rotation test. Knowledge of spatial directions was tested by the question: "You are facing east. Turn right and right again. Where are you looking at now?" (one problem). Estimation of spatial relations and estimation of relative distances in space was tested by asking questions like: "What is closer to the table: the door or the window?" (object-object relation), and "What is closer to you: the TV or the table?" (object-me relation) (2×3 problem). Group test results were compared by Student-t tests and Welch's non-parametric test.

Experimental paradigm and task. The DE was investigated by an experimental paradigm requiring subjects to decide whether the presented number is smaller or larger than five. The distance of the target number from 5 was manipulated. Stimuli were the Arabic digits 1-4 and 6-9. Black stimuli on light yellow background appeared for 800 ms at the centre of a 17 inch computer monitor (800 × 600 pixels) positioned at about 1 m from the subjects' eyes. Trials started with a cross shown for 200 ms. After 1000 ms a digit was shown for 800 ms. The intertrial interval was 2000 ms. 480 stimuli were presented in two blocks, preceded by 72-72 practice stimuli. Responses, counterbalanced across blocks, were given by either the left or right index finger. Reaction times and accuracy were also recorded.

EEG acquisition and processing. EEG was recorded at 33 channels, following the standard electrode sites according to the international 10-20 system: Fp1, Fp2, F9, F7, F3, Fz, F4, F8, F10, Fc5, Fc1, Fc2, Fc6, T9, T7, C3, Cz, C4, T8, T10, Cp5, Cp1, Cp2, Cp6, P9, P7, P3, P4, P8, P10, O1 and O2. The recording was re-referenced to Pz, sampled at 1000 Hz, bandpass-filtered between 0.15-70 Hz, notch-filtered for 50 Hz, than offline-filtered (0.3-30 Hz) and recomputed to average reference. Epochs extended from -100 to 600 ms relative to stimulus presentation (701 time points). The -100 to 0 ms interval served as baseline. Epochs where amplitude exceeded ±100 μV at any of the electrodes and epochs showing ocular artifacts were rejected from analysis. For data analysis and statistics MatLab 7.1 (MathWorks) and Statistica 7 (StatSoft) were used.

Behavioural data analysis. Only response times (RT) belonging to accepted ERP epochs were analysed. Accuracy, raw, and individually normalized RT were examined by Group × Response Side (Left is smaller and Right is larger vs. vice versa) × Laterality (smaller than 5 vs. larger than 5) × Distance (1, 2, 3 and 4) ANOVAs. Normalization served to suppress individual differences and to focus on the pattern of the distance effect. The slope of the DE was calculated by computing the difference between consecutive levels of numerical distance (distance 1 minus distance 2, distance 2 minus distance 3, and distance 3 minus distance 4 values for both levels of Laterality). The stability of the distance effect was assessed individually by subjecting all trials of individual participants to Laterality × Distance ANOVAs in each individual.

EEG data analysis. Trials receiving correct responses within 250-1000 ms relative to stimulus presentation were accepted for behavioural and EEG analysis. Event-related potential (ERP) amplitudes were tested by point-by-point Laterality (1-4 vs. 6-9) × Distance (1, 2, 3 or 4) ANOVAs in each group. Intervals showing significant effects between 100-500 ms for durations longer than 15 data points were considered to demonstrate significant effects. In order to exclude random effects a conservative ($p<0.025$) significance level was used, and only at least partially graded DEs were considered relevant. Partial gradedness means here that the amplitude in conditions Distance 1 and 2 were both either smaller or larger than the amplitude in conditions Distance 3 and 4. Full gradedness means here that amplitudes reflected the pattern of magnitude relations (either Distance 1 < Distance 2 < Distance 3 < Distance 4 or

Distance 1 > Distance 2 > Distance 3 > Distance 4). When point-by-point ANOVAs identified significant effects, the mean voltage at significant electrodes was further subjected to overall Laterality × Distance × Electrode ANOVAs). The topography of the DE was analyzed by a Group × Side (left vs. right) × Location (frontal, fronto-central, central, centro-parietal, parietal and occipital) × Extremity (towards the central line: extreme left, left, central, right, extreme right) ANOVA run on difference waves computed as amplitude for small distance (average of distance 1 and 2) minus large distance (average of distance 3 and 4). Greenhaus-Geisser ε correction was used when necessary. Original F and df, and corrected p values are shown.

2.3. Results

2.3.1. Behavioural tests

Results are shown in Table X/1**.** There was no difference between DG and CG in the Tolouse-Pieron test, forward and backward digit-span. Both groups had ceiling accuracy in TMT-A and B. Nevertheless, TMT-A and B completion time was longer in DG than in CG. DG and CG did not differ in the manipulation of spatial directions, estimating spatial relations and distance. DG and CG did not differ on the body representation task. However, GC outperformed DC in the finger rotation and finger naming tasks, as well as in tactile object identification.

Additionally, although the administration of the IQ tests were separate for the two groups so forth the obtained scores has to be handled with caution, their scores were compared with t-tests. IQ, VQ and PQ all differed significantly in the two groups ($p<0.01$)

						range		Levene		Student-t		Welch-t		
TESTS		**Average**		**stand.dev.**		**(min, max)**		**F(1,14)**	**p**	**F(1,14)**	**p**	**t**	**df**	**p**
		DG	CG	DG	CG	dysc.	cont.							
I	Toulouse-Pieron	96.0	97.1	1.9	3.4	94-99	90-100	0.9	0.34	-0.8	0.45			
	Digit-span forward	6.0	6.6	0.8	1.1	5-7	5-8	1.7	0.21	-1.1	0.3			
	Digit-span backw.	4.7	5.1	0.9	0.9	3-6	4-6	0	0.88	-0.9	0.4			
	TMT-A	25.0	25.0											
	TMT-A time (sec)	88.1	27.3	28.1	9.5	53-133	18-44	6.1	*0.03*			5.4	7.4	**0.02**
	TMT-B	25.6	25.6	1.1	1.1	23-26	23-26							
	TMT-B time (sec)	132.3	57.3	29.3	3.5	97-163	53-63	16.7	*<0.01*			6.7	6.2	**<0.01**
II	Body representation	5.0	5.0											
	Mental hand rotat.	60.9	9.0	11.8	4.1	41-72	67-79	5.4	*0.04*			-3.2	7.5	**0.01**
	M. hand Rotat. Time (sec)	452.4	290.7	64.7	59.9	390-570	225-386	0.1	0.74	4.9	**<0.01**			
	Finger naming	12.6	14.3	1.3	1.5	11- 14	11-15	0	0.93	-2.3	**0.04**			
	Tactile identification	17.3	19.6	2.0	0.8	14-20	19-20	3.3	0.1	-2.8	**0.01**			
III	Shepard's ment. rot. test	23.4	27.3	8.8	4.7	11-36	20-36	3.9	0.07	-1	0.33			
	Directions	1.0	1.0											
	Spatial relations: obj.-obj.	3.0	3.0											
	Spatial relations: obj.-me	3.0	3.0											

Table X/1: Neuropsychological tests.

I. Tests used for measuring general cognitive abilities

1. Tolouse-Pieron test for the measurement of distributed attention, signal detection and concentration
2. Digit span for measuring the STM

II. Tests used for measuring field-specific abilities

1. Egocentric mental rotation (mental rotation based on the representation of the body more precisely on the hands): 4 × 20 picture of a hand rotated by 0º, 45º, 90º or 180º.
2. ‘Allocentric’ mental rotation (Shepard)
3. Trailmaking A and B (Spatial seriality, sensomotor coordination, strategy switching)
4. Tactile-perceptual abilities: stereognosia (object identifying by hands with closed eyes.)
5. Body representation:
    - Autotopognosia. Following instructions e.g. “touch your right ear” “point my (the tester) left eye”, etc. (5).
    - Finger gnosia(?): 15 tasks like naming the touched finger with closed eyes, with opened eyes and showing the requested finger.
6. Spatial operations: e.g. “You are facing East. Turn right and turn right again. Where are you looking at now?”
7. Spatial relations, relative distances:
    - Object-object relation, e.g. “Witch one is closer/farer to the table: the door or the window? ”
    - Object-me relation, e.g. “Witch one is closer/farer to you: the TV or the table?

2.3.2. Magnitude representation: behavioural data

Behavioural results are summarized in Fig. X/1. There was no group effect (RTs /means and SIII/ in adults, CG and DG, respectively: 432±67 ms; 471±79 ms and 503±37 ms. $p>0.14$. Accuracy: 96.7±4.4%; 94.4±6.5% and 92.0±7.3%. $p>0.16$). Response side ($p>0.46$) and the Group×Response Side interaction ($p>0.95$) was not significant. There were more correct solutions in case of larger than smaller distance ($F(3,54)=25.36$; $\varepsilon=0.423$; $p<0.0001$). RTs were shorter to larger than to smaller distance (Raw RT: $F(3,54)=121.63$; $\varepsilon=0.964$; $p<0.0001$. Normalized RT: $F(3,54)=145.42$; $\varepsilon=0.97$; $p<0.0001$). In adults and CG response times were shorter for numbers smaller than 5 than for numbers larger than 5. This pattern was opposite for DG which resulted in a Laterality × Group interaction (see insert in Fig. 1A. Raw: $F(2,18)=4.08$; $p<0.04$. Normalized: $F(2,18)=4.15$; $p<0.04$). There was a Group × Distance interaction in normalized RT ($F(6,54)=4.49$; $\varepsilon=0.8$; $p<0.004$). In order to investigate this interaction, a Group × Laterality × Slope ANOVA was run. The slope of the DE was less steep in DD (0.11 standard deviation) and DG (0.14 standard deviation) than in adults (0.2 standard deviation). Group effect: $F(2,18)=7.05$; $p<0.006$. Tukey contrasts: DD vs. adults: $p=0.0052$. DG vs. adults: $p=0.0415$). Individual RTs are shown in Fig. 1B. The DE was significant in the majority of subjects (All adults: $p<0.0001$. Six out of 7 CG subjects: $0.0001<p<0.002$. Six out of 7 DG subjects: $0.025<p<0.047$). The DE was insignificant in one DG subject ($p>0.3$), and marginally significant ($p<0.08$) in one of the control subjects. The statistical analyses were re-run including all trials (not only the ones accepted for EEG analysis). These secondary analyses yielded statistically identical results to the above described ones.

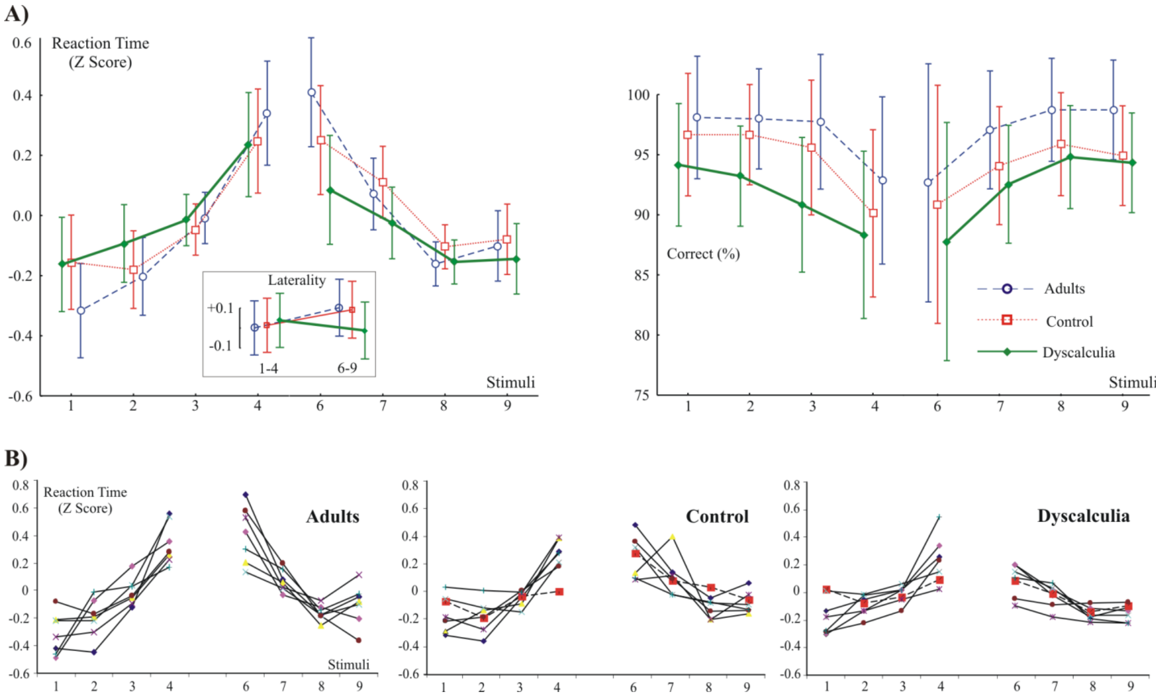


Figure III/1: A: The Distance Effect in the normalised RTs (Z Scores) and in the proportion if corerct answers (%) in all 3 groups. Insertion: the Laterality Effect on the normalised RTs. B: The patterns of the individual Distance Effects Effect in the normalised RTs (Z Scores) in all 3 groups. Individual data marked by dotted line didn't reach significance.

2.3.3. Magnitude representation: ERP data

75-98% of trials were accepted for EEG analysis in individual subjects, i.e. there were around 100 epochs in each Laterality × Distance cell of each subject. ERPs are shown in Fig. X/2. Adults showed a fully graded DE over the left hemisphere between 340-400 ms ($F(1,6)=64.261$; $p<0.001$). The CG showed partial DEs mainly over the right hemisphere between 220-300 ($F(1,6)=65.539$, $p<0.001$) and 400-440 ms ($F(1,6)=7.719$, $p<0.04$). The DG showed partial DEs mainly over the right hemisphere between 210-270 ms ($F(1,6)=7.555$, $p<0.04$) and at some electrodes between 370-430 ms ($F(1,6)=6.251$, $p<0.05$). Topographies of the DE were compared in the above intervals (340-400 ms in adults vs. 400-440 ms in CG vs. 370-430 ms in DG. And: 220-230 ms in CG vs. 210-270 ms in DG). The Group × Side interaction was significant ($F(2,18)=3.98$; $p<0.037$). Post-hoc Tukey contrasts revealed that the topography of the DE differed over the left hemisphere between adults and GC (adults left side vs. CG left side: p<0.016). Further, the topography of the DE differed over the left hemisphere between adults and GD (adults left side vs. DG left side: p<0.037).

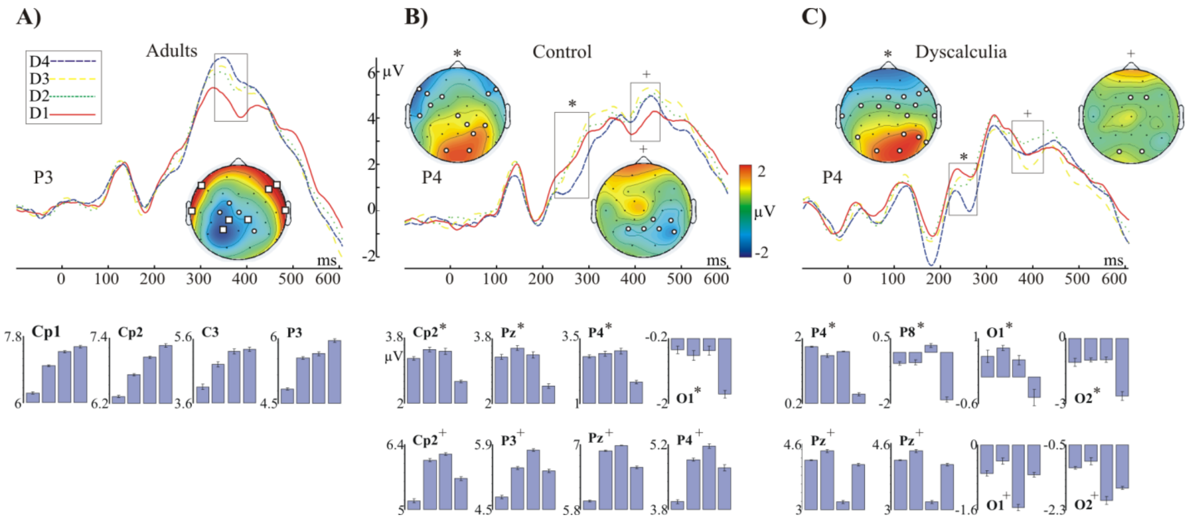


Figure III/2: The raw waves and the topographies of the Distance Effect. Topographies were constructed from de difference between DS and DB. Electrodes with partial Distance Effect are marked by circles. Electrodes with fully graded DE are marked by squares. Point-to-point ANOVA was significant (p<0.025) over these time intervals: A**:** Adults: 300-400 ms. B**:** Control: * 240-300 ms and $^{+}$ 400-440 ms. **C:** Dyscalculia: * 210-300 ms and $^{+}$ 330-430 ms.

2.4. Discussion of the results

The normal behavioural DE in DG suggests that magnitude relations were interpreted by a phenomenologically near normal magnitude representation in the range of single-digit numbers. The slightly faster RTs for larger (6-9) than smaller (1-4) numbers and weaker individual DEs in DG compared to CG indicate that the perception of relative and absolute magnitude of single-digit numbers may have been slightly abnormal in DG. It remains an open question whether larger distortions would be found for larger numbers than the ones used here. The overall topography of early ERP DEs was remarkably similar in DG and in controls. This suggests that early, more automatic processing of digits was similar in both groups. The intact behavioural DE and the similar early ERP DE in both DG and CG is in agreement with a recent fMRI study (Kucian et al., 2006) which found no difference between brain activation of 11-year-old children with DD and control subjects in non-symbolic magnitude comparison (subjects compared the numbers of items in displays). As both symbolic and non-symbolic magnitude comparison relies on the IPS (Dehaene at al., 2004), it seems likely that the magnitude representation of the IPS was close to normal in our DD patients, at least in single-digit number comparison.

The DD subjects differed markedly from the controls in several points. First, in contrast to the similar early ERP DE, late ERP DEs appeared over several right-parietal electrodes in controls, but there was no such focused effect in DG. This suggests that processing differences between DG and controls arose later, during more controlled processing. This view is supported by the accurate but slow TMT performance in DG, which suggests a general deceleration in complex executive functioning. DG subjects did not differ from controls in general mental rotation and in body parts knowledge. In contrast, they were impaired markedly in mental finger rotation, finger knowledge, and tactile performance.

These findings may be interpreted speculatively as follows. First, impaired finger knowledge may be an indication of disrupted angular gyrus function in DD (Rusconi et al., 2005). However, our subjects were not impaired in verbal automatisms and in verbal memory, which might also be expected in angular gyrus dysfunction (Dehaene et al., 2003). This contradiction may be resolved by assuming that the angular gyrus holds more than one representation relevant for arithmetic knowledge (e.g. one of verbal nature (Dehaene et al., 2003) and another related to finger representation (Rusconi et al., 2005)). This assumption is supported by an intracranial electro-stimulation study which identified areas within the left angular gyrus which were specifically involved in either multiplication or subtraction (Duffau et al., 2002). We may also speculate that the deceleration of executive functioning makes it more difficult to focus attention on numerical information, thereby slowing performance in more complex arithmetic tasks (Dehaene et al., 2003). This may also explain why Kucian et al. (2006) found fMRI activation differences between children with DD and control subjects in an approximation task, but not in non-symbolic magnitude comparison nor in exact calculation. Their approximation paradigm involved selecting a close (but never perfect) solution to arithmetic problems. This task may tax the network directing attention to magnitude more than the other two tasks used by Kucian et al. (2006). It is unclear whether the slight abnormality in the perception of magnitude, the impairment of executive functioning, and the impairment of finger knowledge were all due to the disorder of one mental representation relevant for arithmetic performance. It seems to be more likely that the impairment of more than one relevant representation contributed to decreased performance in DD. Further, preferably longitudinal, studies should investigate whether the impairment of a certain mental representation form precedes (and causes) the impairment of other mental representations relevant for arithmetic performance in DD.

Our data also suggest that the neural networks underlying or the cognitive strategies used during magnitude discrimination performance may undergo development even during adolescence. First, the slope of the DE was less sharp in adolescents than in adults. Second, the direction of the ERP DE was reversed in both adolescent groups. Third, the topography of the late DE differed markedly both between adults and adolescents. Fourth, a fully graded ERP distance effect appeared only in adults.

3. An electro-physiological temporal principal component analysis of processing stages in number comparison and developmental dyscalculia[14]

Abstract

By a re-analysis of a previous event-related brain potential (ERP) data [section 2] here my objective was to identify and compare cognitive processes in adolescents with DD and in matched control participants during one-digit number comparison. To this end temporal principal component analysis (tPCA) was performed on ERP data. First, tPCA has identified four major components explaining the 85.8% of the variance in number comparison. Second, the ERP correlate of the most frequently used marker of the so-called magnitude representation, the numerical distance effect, was intact in DD during all processing stages identified by PCA. Third, hemispheric differences in the first temporal component and group differences in the second temporal component suggest executive control differences between DD and controls.

3.1. Rationale and background

The aim of the present study was to gain a clearer insight into the cognitive processes contributing to the variance of ERP data recorded previously (section 2). With the reanalysis of the previous data my main objective was to define more clearly what temporal stages were involved in number comparison in normal adolescents and in adolescents with DD. Temporal principal component analysis (tPCA) method was used on the ERP data. tPCA decomposes ERP into independent principal components in time. Hence, tPCA is able to delineate cognitive processing into (ideally) independent stages by discriminating amongst temporally overlying, but functionally distinct components. This is especially important for ERP data which consists of the superimposition of ERP correlates of several overlapping processing stages.

---

[14] This Chapter, with some modifications, has been published: Soltész and Szűcs, 2009, Cog. Dev.

ERPs in CG and in DG for distance 1 (D1) and distance 4 (D4) are shown in Figure III/3. For further details of ERP analysis and results, see section 2.

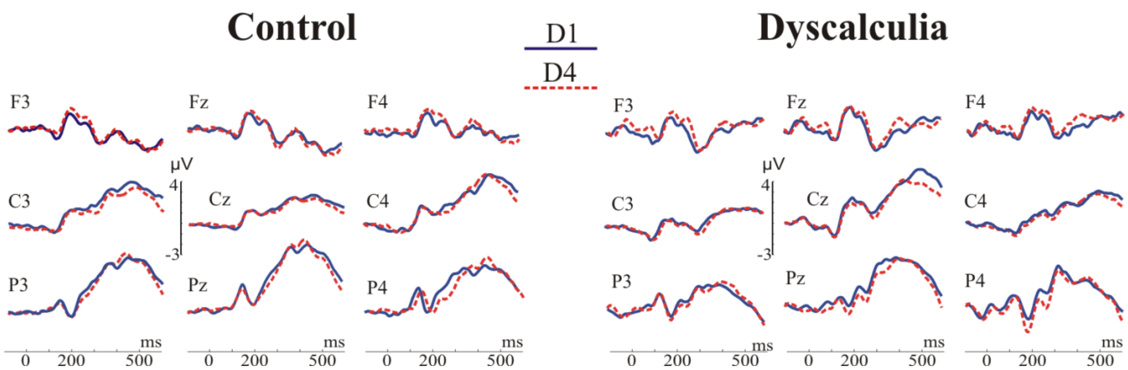


Figure III/3: ERPs for CG and DG in the two experimental conditions at frontal, central and parietal electrodes.

3.2. Methods: Temporal principal components analysis (tPCA) procedure

Detailed description of PCA calculations are presented in Chapter II/3. The data matrix of time points [701], electrodes [32], subjects [14] and conditions [2] was first transformed into a matrix where columns represent variables (time points) and rows represent observations (electrodes, subjects and conditions). Preceding tPCA, ERPs were down-sampled to 250 Hz by averaging consecutive time points in a 4 ms window, in order to enhance signal-to-noise ratio and to lower the number of variables in comparison to observations to enhance the power of PCA (Chapman and McCrary, 1995). Covariance matrix of the variables was calculated and entered into PCA. PCA was performed by a custom written toolbox using MatLab 7.1 SVD algorithm for singular value decomposition. Temporal components to retain were selected using the Scree test (Cattell, 1966).

3.3. Results

Seven components accounting for the 96.7% of the variance in the original data were retained. Factor loadings for the seven principal temporal components (TC) are shown in Figure III/4, the variance accounted for by the individual TCs after varimax rotation are: 62.59%, 8.41%, 11.21%, 7.44%, 2.07% 3.59% and 1.44%, respectively.

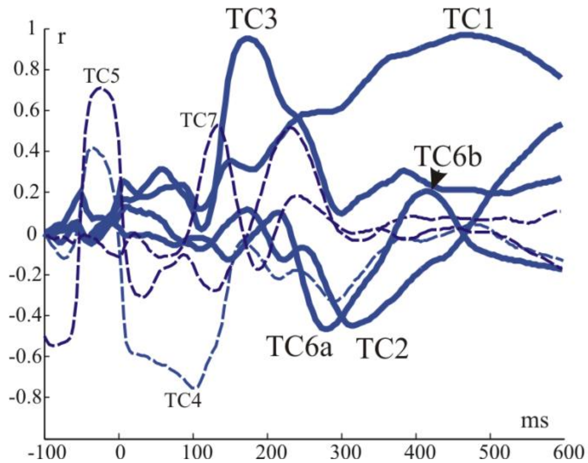


Figure III/4
Component loadings of the 7 selected components. Loadings represent correlations between the given component and the variables.

The signs (+ or –) of component loadings are arbitrary according to the SVD computations. Component loadings were multiplied by -1 if the sign of the absolute maximum value of corresponding ERP with the component's time interval and topographic distribution was the opposite to that of the component's sign. Again, factor loadings here are the correlations between the factor and the original variables and depict the time courses of the factors. High correlation between a factor and a time point means that the given factor contributes largely to the given time point. It is important to note that components with broader distribution in time tend to take up more variance than components with narrower distribution, because broader componets include more time points co-varying with each other, therefore explaining more variance. Nevertheless, they are not necessarily more important or meaningful than components explaining less variance – it is the variance in the factor scores what really matters (Boxtel, 1998). Factor scores are the values along each retained factor, or virtual ERP, in the new factor space. The calculated factor scores are estimations of the values we had gained for each principal component at each electrode in each condition and subject, if we had measured these principal temporal components only instead of the original time points.

The time course of the loading of TC1 has a slow onset at around 300 ms and loads the most at around 400-500 ms. Factor scores along this vector (TC1) have a positive peak at centro-parietal electrodes (see Figure III/5).

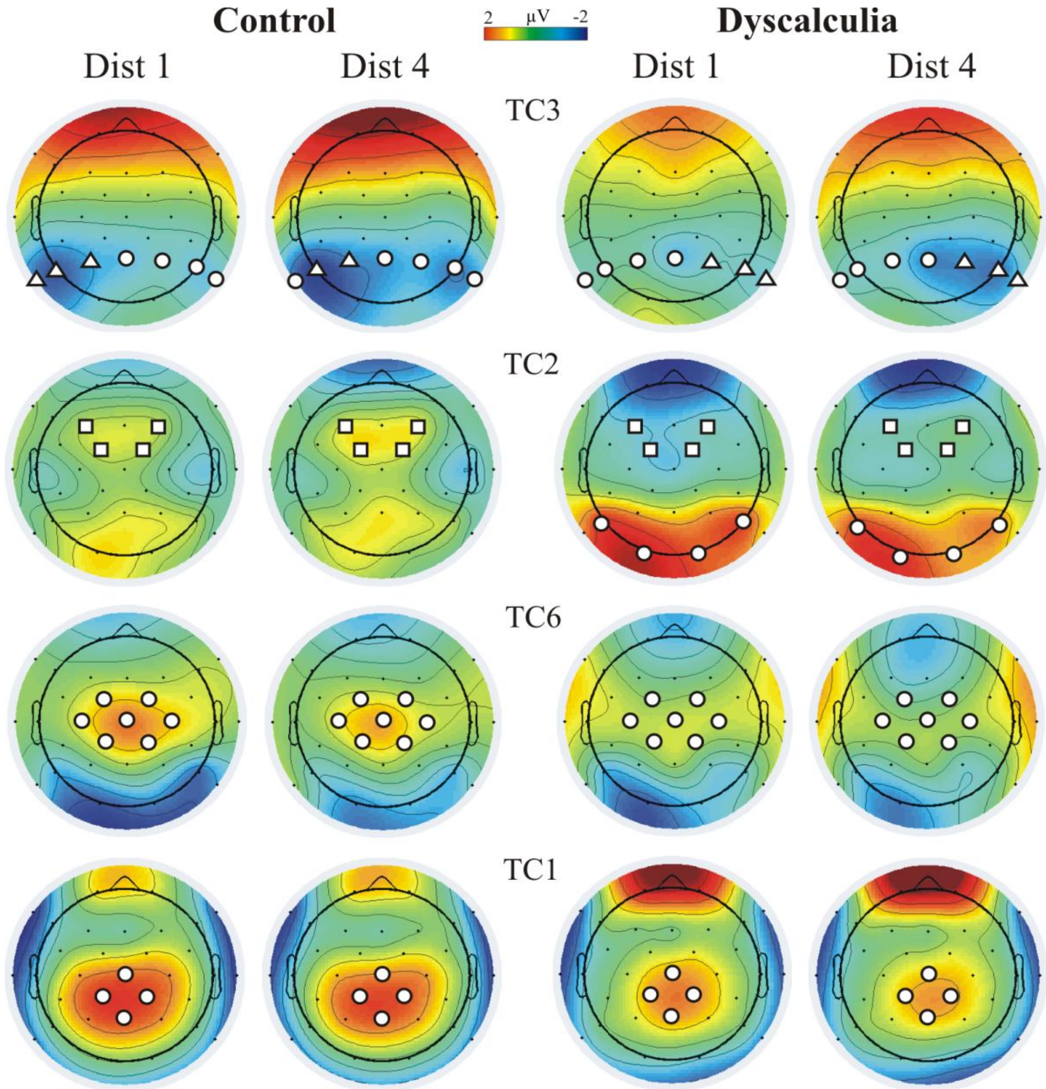


Figure III/5: Component scores for TC3, TC2, TC6 and TC1. The represented values are in untits of microvolts. TC3: markers denote electrodes with significant DE. Triangle markers show the significant hemispheric differences. TC2: circle markers denote significant DEs. Square markers denote group differences in polarity, where DE was also significant. TC6. Markers denote significant DE. TC1: markers denote electrodes at wich DE was significant.

The time pattern of TC1 is tipical to the P3b ERP component, just as the topographical distribution of the factor scores of TC1. TC2 has high negative loadings around 300 ms (TC2a) and high positive loadings starting at around 500 ms (TC2b; unfortunatelly, the time window of -100 to 600 ms does not allow us to see the full time scale of this component). The topographical distribution of TC2a scores indicates a negative peak (in DG only) above fronto-central locations and a positive counterpart

(TC2b) at occipito-temporal areas. The timing and topographical distribution of TC2a is akin to the so-called N300 ERP component described in language and picture naming experiments and is found to be sensitive for semantic mismatch in the stimuli (Hamm et al., 2002, Federmeier and Kutas, 2002). TC3 loads at 200 ms and has a negative occipito-parietal distribution, reminiscent to a member of the N2 family. A parietal negativity around 200 ms reflects the focusing of spatial attention (Luck and Hillyard, 1994, Luck et. al., 1997, Eimer, 1996), and is also linked to some aspects of working memory maintenance (Vogel and Machizawa, 2004). The biphasic TC6 loads negatively at around 280 ms (TC6a, see Figure III/4) over occipito-temporal areas and has a positive peak around 400 ms over cental locations (TC6b). Functional meaning and relevance is unclear of this component, but because of its timing partly overlaps with the previous components TC6 was also analysed.

The remainder three TCs are excluded from further analyses. TC4, TC5 are most probably reflect autocorrelation of the data due to baseline correction (Boxtel, 1998) and noise from eye movements, and the early biphasic TC7 may represent early visual components.

Temporal components are discussed in the order of their timing and not in the order of their explained variance. As it was menioned before, the magnitudes of explained variances does not set up an order of importance in TCs.

TC3. The parietal electrodes Pz, P3, P4, P7, P8, P9 and P10 were subjected to an ANOVA with Distance and Electrode as within-subject factors and with Group as between-subject factor. A significant Distance effect and a significant Electrode × Distance interaction were found (Distance: $F(1,12)=24.87$, $p<0.001$; Electrode × Distance: $F(6,72)=7.22$, $p<0.001$). Because of the striking side differences on the topographic maps, the electrodes were re-organized into a new structure with the statistical factors Hemisphere (left and right) and Location (close to midline, far from the midline and at the brim) and were re-entered into a Group × Distance × Hemisphere × Location ANOVA. The interaction of Distance × Hemisphere × Location × Group was significant ($F(2,24)=7.28$, $p<0.01$). Post-hoc comparisons showed that the two hemispheres differed in both distances ($p<0.001$). In CG, factor scores at left electrodes (P3, P7 and P9) were more negative than the scores at the right electrodes (P4, P8, P10) (see Figure III/5 and Figure III/6A and insert in Figure III/6A). Similarly, the Hemisphere difference was significant across all Locations in DG, however, with the oposite direction: factor scores of the right electrodes were more negative than factor scores of the left electrodes.

TC2. The fronto-central F3, Fc1, Fz, F4 and Fc2 elesctrodes (TC2a) and the occipito-parietal O1, P7, O2 and P8 electrodes (TC2b) were entered into an ANOVA with Distance and Electrode as within-subject factors and with Group as between-subject factor. At the fronto-central electrodes, a significant Distance effect and a significant Group effect were found (Distance: $F(1,12)=18.72$, $p=0.001$; Group: $F(1,12)=10.64$, $p<0.01$). The significant group effect reflects the opposite polarity of the component in CG and DG (see Figure III/6 and Figure III/6B). At occipito-parietal sites a significant Distance effect a significant Group effect and a significant Electrode × Distance interaction were found (Distance: $F(1,12)=4.57$, $p=0.05$; Group: $F(1,12)=10.64$, $p<0.01$; and Electrode × Distance: $F(3,36)=2.93$, $p<0.05$). Post-hoc analyses revealed that the distance effect was significant only in DG at O1, O2, P7 and P8 ($p<0.01$ at O1, O2 and P8; $p<0.06$ at P7) while it was not significant in CG at any of these electrodes (see Figure III/5 and Figure III/6C).

TC6. The occipito-temporal O1, P7, O2 and P8 (TC6a) and the central electrodes Fc1, C3, Cz, Fc2, C4 and Cp2 (TC6b) electrodes were entered into an ANOVA with Distance and Electrode as within-subject factors and with Group as between-subject factor. At the central electrode sites, there was a significant Distance effect ($F(1,12)=8.04$, $p<0.02$, see Figure III/5 and Figure III/6D). No Group effect or interaction including Group was significant. Statistical analysis of TC6a scores did not yield any significant results.

TC1. Based on the topographical distribution of TC1, temporal scores of centro-parietal electrodes Cp1, Cz, Pz, Cp2 and of frontal electrodes Fp1 and Fp2 were subjected to an ANOVA with Distance and Electrode as within-subject factors and with Group as between-subject factor. A significant Distance effect was found (Distance: $F(1,12)=6.38$, $p<0.03$, see Figure III/5 and Figure III/6E). There were no significant effects of any sort on the two frontal electrodes. There were no statistically significant differences between the two groups and there were no significant interaction involving the Group factor.

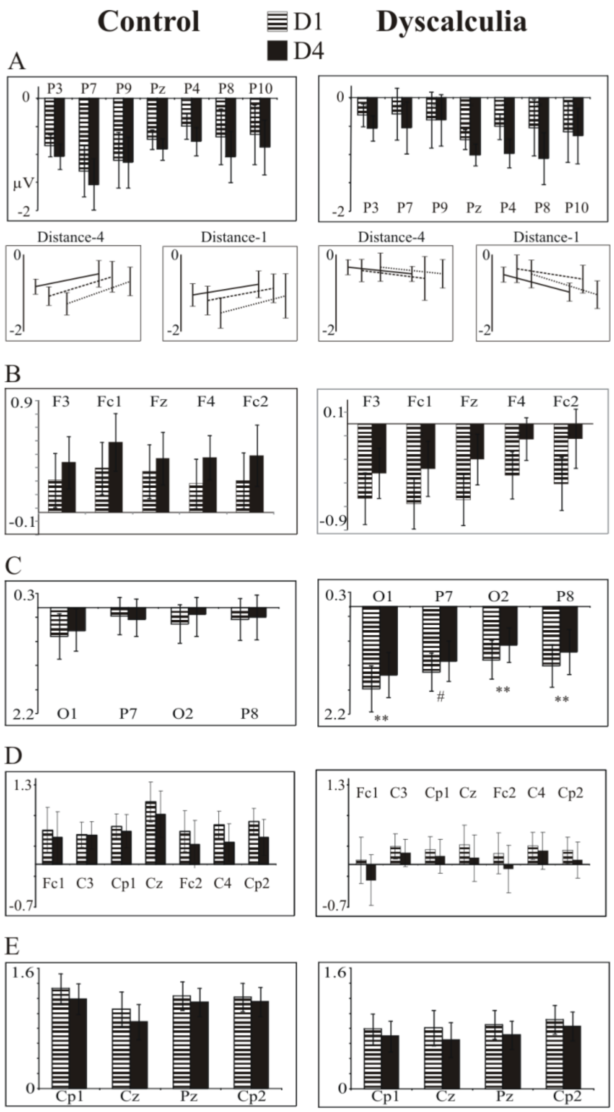


Figure III/6: Temporal component scores for the two distances in DG and CG, at the relevant electrode sites. Units are in microvolt values, representing the difference from the grand mean of the data. A: Component scores for the two distances are represented for both groups for TC3 at the relevant electrodes. B: Component scores for TC2 at the frontal electrode sites. C: Component scores for TC2 at the occipital electrode sites. (**) denotes significant distance effect at p<0.01. (#) denotes distance effect at p<0.06. D: Components scores for TC6 at fronto-central locations. E: Component scores for TC1 at centro-parietal locations.

3.4. Discussion of the results

PCA decomposition of the averaged waveforms complements the previous results (Section 2) by explicitly defining independent processing stages explaining most of the variance (85.8%) of the data. Further investigations of the processing stages identified here may shed light on functional differences between dyscalculics and control subjects.

On the one hand, the present results demonstrated that the DE was significant in principal components in DD adolescents and was not dissimilar from that of controls during all processing stages identified by the PCA. On the other hand, group differences were found in some of the processing stages most probably linked to attentional and control functions. First of all, a hemispheric asymmetry was found between DG and CG in a temporal component peaking at 200 ms with a parietal topographic distribution (component TC3), even though the DE was significant in both groups. This hemispheric asymmetry during the time interval of the distance effect suggests that partially different neuronal networks may be involved in number comparison in DD adolescents and controls. This conclusion is in line with Shalev et al. (1995) who showed hemispheric dysfunction is related to calculation difficulties. The significant DE in our data suggests that the automaticity of number processing and the functionality of the magnitude representation are preserved in DD. This suggests that neural differences do not necessarily affect the magnitude representation, but rather, another part of the processing network. With its timing and distribution, TC3 is reminiscent of a member of the N200 family. A parietal negativity around 200 ms reflects the focusing of spatial attention (Luck and Hillyard, 1994, Eimer, 1996), and is also linked to some aspects of working memory maintenance (Vogel and Machizawa, 2004). As neither spatial attention nor working memory was addressed in this ERP experimental paradigm, I cannot draw strong conclusions about the function of this particular component. However, it seems likely that differences in more general executive functions cause the hemispheric dissimilarity between the two groups, and that DD adolescents may rely on different attention-related or working memory related neural networks to solve the task than control adolescents do.

Second, a robust group difference was found in the following temporal component. TC2 had completely different properties in the two groups. In fact, TC2 seems to be present only in the dyscalculia group. The timing and topographical distribution of TC2 is akin to the so-called N300 ERP component described in language and picture naming experiments and is found to be sensitive for semantic mismatch in the stimuli (Hamm et al., 2002, Federmeier and Kutas, 2002). The same component is also linked to action control (Goodale and Milner, 1992), and was localized in the cingulate cortex via source localization and co-registered ERP and fMRI (Vogt et al., 2006, O'Hare et al., 2008). The cingulate cortex is involved in self-monitoring, awareness of and reaction to threatening stimuli; all of these functions can be considered as highly relevant in number processing in DD (usually DD subjects are very anxious when carrying out mathematical tasks). Considering these results together with the fact that DD and control differed in speed but not in accuracy in the Trail Making Test (see section 2), we hypothesize that DD and control differed in problem solving strategies and in executive functioning.

The positive central TC6 appearing at around 400 ms and the late positive TC1 above more posterior sites can tentatively be identified as different members of the P300 family. The time course of TC1 is typical to the P3b ERP component, just as the topographical distribution of the factor scores of TC1. P300 reflects the termination of cognitive processing (Kutas et al., 1977, McCarthy and Donchin, 1981) and is found to be modulated by numerical distance (Grune et al., 1993, Dehaene, 1996). P300 is often related to a mechanism for updating contextual information in working memory (Donchin, 1981, Donchin and Coles, 1988). However, the 'family' of P300 has been decomposed by PCA into at least three different components, and they were showed to be sensitive to different experimental manipulations, suggesting functional differences among them (P300a, P300b and Slow Wave; Spencer et al., 1999, 2001). Because no other manipulations than numerical distance were introduced in this particular experiment, we cannot differentiate the two components in functional terms. What I can suggest based on the data is that the temporal component around 400 ms and a slow late positivity in the 200-600 ms time window may reflect task-relevant decisional processes and probably non-specific motor preparation as well (Dehaene, 1996), which are both influenced by numerical distance.

Overall, the present data suggests that the timing of magnitude processing and magnitude representation is functionally intact in DD. At the same time impairments in other cognitive functions, especially in executive functions, may culminate in prominent difficulties with arithmetics. Executive processes have been shown to play a major role in the calculation network. For example, Menon, Rivera, White et al. (2000) found that accurately calculating normal adults showed less frontal activation during arithmetic than less accurate calculators. This finding has been attributed to more automatized calculation processes in better calculators relative to worse calculators. Similarly, children may show a frontal to parietal shift during development which may be related to their less automatic arithmetic skills than adults' (Ansari et al., 2005; Rivera et al., 2005). Hence, deficient executive functions can result in difficulties during arithmetic development. Alternatively, one could argue that deficient number representation leads to differences in executive functioning during numerical tasks, by placing more demand on control functions and other problem solving strategies in order to compensate for the weakness of the magnitude representation. In contrast to this alternative hypothesis, we not only found practically normal magnitude processing in DD by both ERPs and PCA, but DD and control participants also differed significantly on tests measuring executive functioning (reported in detail in section 2). Naturally, the present study has its limitations due to the relatively low number of DD participants and because it has been shown that selection criteria and tests used for selection may include or exclude different types of DD (Mazzocco & Myers, 2003, Murphy et al., 2007), yielding results which does not allow generalization of the results to the whole DD population (however, the heterogeneity of DD already implies that there may be more than one explanatory factors behind DD).

4. Developmental dyscalculia: general discussion

In both the ERP and tPCA data significant DEs were found in subjects with dyscalculia. Moreover, DEs appeared with a similar timing in both controls and in subjects with dyscalculia in all processing stages identified by tPCA. This suggests that the automatic processing of numerical information occurs with similar speed in both DD and control adolescents. These findings indicate that not all cases of DD are due to the dysfunction of the magnitude representation. Rather, a temporal component with a fronto-central negative distribution followed by a later occipito-temporal positivity, both of which are not present in the control group, suggest monitoring function differences in DD.

4.1. The reconciliation of domain-specific and domain-general hypotheses

In summary, the deficit of the core magnitude representation could not be supported by either the present study (or by many of the previously reviewed ones). Rather, in my opinion it is more reasonable to argue that slight impairments in the working memory system can lead to more prominent difficulties in mathematics.

However, we cannot exclude the possibility of a 'pure DD' caused merely by the deficit of the core magnitude representation. For example, Rubinsten and Henik (2009) and von Aster and Shalev (2007) propose that both the core magnitude representation and working memory are indispensable for numerical development. If either of these systems does not function properly, difficulties in mathematics will unavoidable appear. These authors suggest drawing a distinction between *developmental dyscalculia* and *mathematical learning disabilities.* The former is the result of a deficient core magnitude representation, while the latter is the result of the interplay of deficits in other abilities, like language or working memory.

While this proposal sounds reasonable, unfortunately there is no such a report in the peer-reviewed literature which would present a subject with DD, who shows an isolated deficiency or absence of the magnitude representation[15].

4.2. Difficulties with building up and accessing numerical representations from symbols

Deficits of the working memory are supposed to contribute considerably to difficulties during the building up of numerical representations, to difficulties in learning and applying calculation rules and understanding the underlying concepts of mathematics (Geary, 1993). It is easy to see that an impaired working memory system leads to the formation of a weaker than normal representation of symbols and arithmetic facts in the long term memory; these weaker representations, together with the difficulty of keeping and manipulating information in the short term memory can cause prominent problems during the acquisition of mathematics.

Supporting this hypothesis, Rouselle and Noël (2007) reported that children with DD did not show any deviations from normal in the numerical distance effect during non-symbolic number comparison; the distance effect was found to be weaker during symbolic number comparison only. Further, although showing a pattern of congruency effects reminiscent of younger age groups' in the numerical Stroop paradigm (see for example Girelli et al., 2000; Rubinsten et al., 2002), the magnitude representation was accessed automatically in DD children as well. Rouselle and Noël conclude that it is not the basic 'magnitude representation' which is deficient in DD, but rather the process of accessing these representations from symbols.

[15] Butterworth (1999) reported one case in his book, Charles, a 30-year-old psychology graduate student who showed a reversed distance effect suggesting an abnormal magnitude representation. Unfortunately, this case was not reported in any peer-reviewed journal.

4.3. The neuroconstructivist perspective

There is one more line of argument which is unfortunately neglected by many in number research. Karmiloff-Smith (1995; 2006; 2009) and Ansari and Karmiloff-Smith (2002) apply a different approach to development and to developmental disorders. They claim that neuro-cognitive development departures from a state in which cognitive functions are neither localized nor very specialized at birth. During development, the small building blocks of cognition all contribute to the learning and to the specialization of higher-level functions. The slightest and even unrecognized abnormality in one of these building blocks can initiate a cascade of devious developmental routes, possibly resulting in a major and easily recognizable problem in more complex functions, like language or mathematics. Karmiloff-Smith (2006) also notes that even if the IQ is in the 'normal range' in some these specific groups with developmental cognitive disorders, there might be significant differences in IQ. For example, subjects with specific language impairment usually have an IQ score that is still considerably lower than the average. In DD research, most people use the criterion of >85 scores for the diagnosis of DD. Clearly, an IQ of 89-99, even if being in the normal range, is not comparable to the average of the normal population, which is around 110 scores. Furthermore, the mean IQ score of the seven DD participants in my study was also significantly lower than that of the control groups. In fact, most of the studies either do not report IQ values, or do not contrast IQ values between groups.

4.4. Conclusion

After all, DD seems to be a very heterogeneous disorder with multiple possible underlying deficits in the cognitive system. In accord with this variability, there is no consensus on the *diagnosis criteria* of DD yet, let alone the proposed remediation methods for special education. In the future, more emphasis on the investigation of *non-numerical* cognitive abilities is needed. Of course, the possibility of a domain-specific impairment to the magnitude representation can not be excluded in DD, but narrowing down the research to the supposed deficit of this magnitude representation yields misleading conclusions and badly-designed remediation plans for DD. In my opinion, mathematics is such a complex domain which can not be explained in simple terms. If one has slight deficits in for example working memory and attention, these deficits may not lead to recognizable problems on the playground, or during reading; but may culminate into apparent difficulties when one has to learn mathematics. Certain types or constellations of weaknesses may not be apparent at all, as long as these weaker than normal systems are not taxed extensively. For example, if one has somewhat weaker than normal knee joints can still be a great swimmer; his weakness would never come to exist as long as he does not try to run the Marathon. Something similar would apply to mathematics as well: a little weaker than normal working memory does not impede the learning of reading, but when it comes to the mental rotation of triangles or to carry and borrow in mind while adding up two numbers, then one might find oneself in trouble.

I suggest that a research approach based on the *neuroconstructivist perspective* would be fruitful for dyscalculia research. Longitudinal examination of several basic cognitive abilities and their interplay during development would provide a more detailed and clearer picture on cognitive development, including the learning of mathematics.

IV. Magnitude representation in children – what counts?[16]

Abstract

There is a debate in the developmental literature whether the analogue magnitude representation preempts and predicts counting (and later arithmetic) abilities or it is rather the other way around: learning to count helps children in the understanding of the abstract nature of numbers, and to understand that numbers are independent of other perceptual magnitudes. May seem as a theoretical question, it is indeed highly practical and has considerable consequences on how counting and mathematics could and should be taught. First I review the current literature on the magnitude representation of children and infants. Second, I critically review the most referred study in the field, which aims to support the hypothesis that (approximate) numbers are abstracted and recognized by infants. Third, a large-scale study is presented, where magnitude representation and counting abilities, among other non-numerical abilities, were tested and correlated with each other. Lastly, the findings are discussed; language and counting knowledge does indeed support the abstraction of numbers.

1. Approximate number representation in children: an overview

It is a major question whether the representation of approximate numerical magnitudes in children develops and sharpens independently of symbolic arithmetical abilities, or symbolic knowledge correlates with the approximate magnitude representation in some ways. There is a sharp divide in the corresponding developmental literature: many argue that the innate magnitude representation is a prerequisite of the acquisition of arithmetics; others claim that formal education and numerical enculturation sharpens the magnitude representation in children. On the one

[16] This chapter, in a somewhat different form, has been reviewed, revised and waiting for editorial decision at Behavioral and Brain Functions.

hand, several researchers assume that children have an innate, preverbal approximate, language-independent magnitude representation shared with other species (Brannon and Terrace, 1998, Gallistel and Gelman, 1992, Dehaene, 1997, Xu and Spelke, 2000, Huntley-Fenner and Cannon, 2000, Brannon et al., 2004, Xu, 2003, Wynn, 1992, Feigenson et al., 2004, Barth et al., 2005a,b). According to this account, refinement of the analogue magnitude representation correlates with math achievement (Booth and Siegler, 2006, 2008, Siegler and Booth, 2004) and has a predictive value for later math performance (De Smedt et al., 2009a, Halberda and Feigenson, 2008). On the other hand, others think that the relation is reversed. Development and sharpening of magnitude representation is supported by language, especially by counting skills (Brannon and Van de Walle, 2001, Rouselle et al, 2004, Rouselle and Noël, 2008).

Number representation skills are usually tested by quantity discrimination tasks. The most general finding is that quantity discrimination performance depends on the ratio of to-be-compared quantities: it is harder to compare quantities if their ratio is closer to 1 relative to when their ratio is further away from 1. As it is already discussed in Chapter I, **t**he ***ratio effect*** has been shown in animals, human infants and human adults. Hence, it is thought that magnitude and number is coded in an analogous, approximate fashion by an evolutionarily grounded pre-verbal magnitude representation (Deheane, 1997; Gallistel and Gelman, 2000; Meck and Church, 1983; Platt and Johnson, 1971; Moyer and Landauer, 1967). The ratio-sensitivity of magnitude comparison results in two effects: The distance effect; it is harder to compare closer than further away numbers (e.g. comparing 6:8 is harder than 5:9). And the size effect; it is harder to compare larger than smaller number pairs if their difference is fix (E.g. 8:12 is harder than 4:8 where (distance is fixed at 4; and: 8:12 is harder than 2:3 where ratio is fixed at 2:3).

Number representation in infants and young children is usually examined in non-symbolic comparison tasks. In these tasks infants are expected and older children are explicitly asked to discriminate between displays showing a certain number of items (e.g. dots). The most important methodological challenge in these experiments is that perceptual variables are correlated with number (Clearfield and Mix, 1999, 2001; Mix et al., 2002; Feigenson et al, 2002). After a thoughtful reconsideration of some of the earlier studies (Mix et al., 2002) it was shown that the problem with some of the paradigms is that these paradigms did not control for some perceptual variables which are in correlation with number. For example, when using the same dots or dolls (e.g. Starkey and Cooper, 1980, Starkey et al. 1983, Wynn, 1992), the overall surface or the overall circumference (the summation of the occupied area of each items, or the summation of the circumference of each items) of the dots is proportionally larger in case of three dots than in case of two dots. Comparison of perceptual features and that of numerical magnitudes leads to the same conclusion: three is more than two. When these perceptual variables are kept constant through number or pitted against number, results showed a more disappointing picture of children's numerical abilities and would support a *continuous quantity account*: children are sensitive to continuous perceptual features instead of number. When overall surface (Feigenson et al, 2002) or circumference (Clearfield and Mix, 1999) is controlled during experiments, infants are more sensitive to the continuous perceptual variable than to number. It was also shown that infants habituated to total surface area but not to number when these two dimensions were pitted against each other, i.e. when the numerically 'more' set was smaller in physical size (Feigenson et al, 2002a).

Xu and Spelke (2000) elaborated an experimental design, in which they tried to rigorously control for continuous perceptual variables. The authors tested 6-month-old infants' quantity discrimination skills with a visual habituation procedure. Continuous variables, like overall surface and density were varied throughout habituation trials and only number remained constant. Conversely, all non-numerical variables that varied through habituation trials were held constant across test trials. Under these well-controlled conditions infants discriminated between 8 vs. 16 dots but not between 8 vs. 12 dots. These results have been replicated and extended by several later experiments (for a review see Feigenson et al. 2004). It has been concluded that even can discriminate between numerosities if they differ in a ratio of at least 1:2, even when confounding variables are controlled for. It has also been found that discrimination performance in infants and children decreases as the ratio of the to-be-discriminated numerosities is approaching one (Symbolic stimuli: Sekuler and Mierkiewitz, 1977; Temple and Posner, 1998. Non-symbolic stimuli: Starkey and Cooper, 1995; Huntley-

Fenner and Cannon, 2000; Huntley-Fenner, 2001; Xu and Spelke, 2000; Barth et al., 2003, 2005a,b; Rouselle et al., 2004; Jordan and Brannon, 2006c; Halberda and Feigenson, 2008).

1.1. Methodological issues

There is an ambiguity inherent to most of the experimental designs and stimuli in non-symbolic number comparison, which needs resolution. Number comparison and discrimination follows the same law as the discrimination of values along other continuous perceptual dimensions: the larger the difference between two objects (for instance between two sizes, luminance values, and numbers), the easier and faster to discriminate between them (Weber's law, e.g. Mechner, 1958; Dehaene, 1997; Brannon et al., 2006). The question can be raised whether the effects assigned to innate number representation (e.g. the ratio effect) could be accounted for by mere perceptual variables? The perceptual variables correlating with number are interdependent with each other and cannot be controlled for at the same time. For instance, if intensive properties (individual item properties, like item size) are kept equal, extensive properties (properties of the item group, like the sum surface of the dots) will inevitably co-vary with number. Whichever perceptual variable is controlled for in one trial, there is always another perceptual variable which co-varies with the number. For instance, if collections with both more and fewer elements contain elements with the same average item size it is easy to see that overall surface of those collections will be different, systematically and proportionally with their numerosity. On the flip side, if overall surface is equated across collections, average item size will systematically co-vary with the collections' numerosity (for an illustration see Figure IV/1).

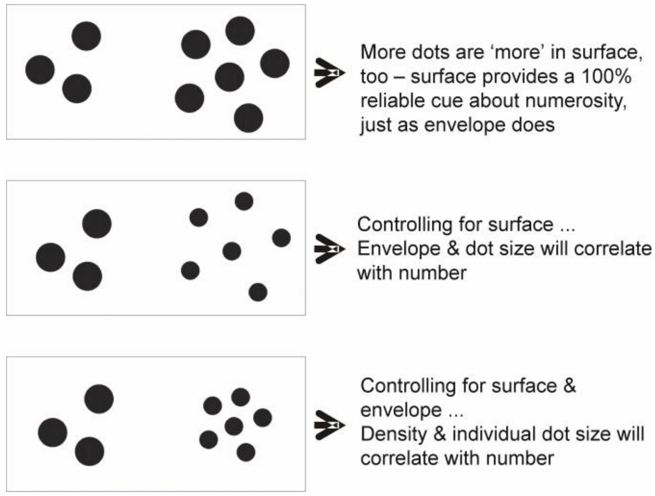


Figure IV/1: Three examples of the co-variation among physical properties and number. Envelope: the outline of the set of objects

So as it is impossible to control for all perceptual variables at the same time within one display, it is very difficult to discriminate between the perception of approximate numerosity and the perception of continuous perceptual variables. The continuous quantity account would postulate that magnitude comparison relies on continuous perceptual variables rather than on number per se.

1.1.1. A critical analysis of Xu and Spelke's (2000) experimental design

As Xu and Spelke's paper (2000) is the most often cited study for the support of the innate approximate magnitude representation, so I think it is more than necessary to have a critical look on their experimental design. Xu and Spelke (2000) attempted to *control* for non-numerical perceptual variables in the following way: they varied the sum surface of the displays via changing item dot size from display to display. Number was held constant, so that number was the only feature of the habituation sequence which did not change – so forth expected to result in habituation to the given number. The test display was designed to have a sum surface of dots falling into the middle of the distribution of the range of sum surfaces during habituation displays (Figure IV/2).

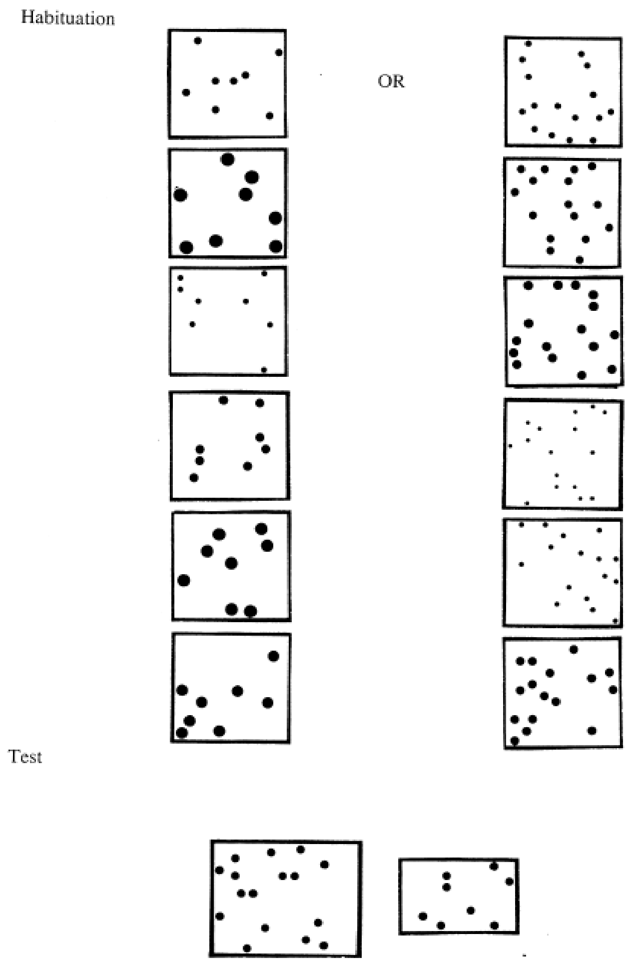


Figure IV/2: Stimuli from Xu and Spelke (2000).

Though, by having a closer look at the stimuli parameters of this design, it can be seen that by controlling for overall surface, item size will inevitably co-vary with number. According to their stimuli description, the diameter of an item varies between 1.06-2.37 cm in the 8-item displays and between 0.75-1.67 cm in the 16-item displays. As item diameter on test displays is 1.5 cm, item size of dots in a 8-item test display is larger than the average item size in a 16-item habituation display (1.3 cm) and larger than the 81% (!) of the item size variations of 16-item habituation displays. On the other side, item size of dots in a test 16-item display is smaller than the average item size of the habituation 8-item displays (1.83 cm), so forth smaller than the 66% of the item size variations of the 8-item habituation displays.

The same confound can be shown for the sum circumference (see also Mix et al., 2002): sum circumference in fact varies proportionally with numerical change. And similar arguments apply to all the other studies which followed this design (for example in Xu et al. 2005, etc.). Experiments conducted in the auditory modality (Lipton and Spelke, 2003) are also subject to perceptual confounds: overall duration, rate, and item duration cannot be manipulated independently of each other.

Regarding the analysis, data presentation and interpretation in the Xu and Spelke (2000) study there are some other issues which can also be criticised. (1) The authors report that the looking time of infants was significantly longer for the test display with novel number of dots, than for test displays which contained the same number of dots than habituation displays. However, they did not report at all, whether the *habituation* was significant at all, or not; they did not test the fall of looking time during habituation trials. (2) 12 of 16 children looked longer at the new number, after 5 other children were excluded; actually, only approximately half of the children looked longer at the new number, than at the old number. (3) The looking time difference between the old and new numbers was significant only in the 2nd trial pair, but not in the 1st or in the 3rd trial pair (see Figure IV/3).

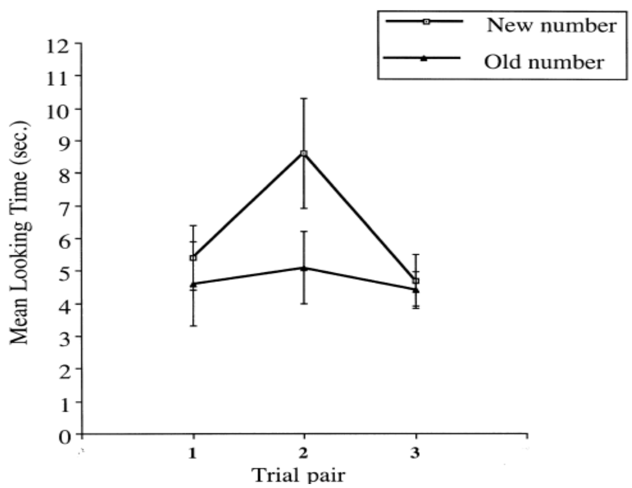


Figure IV/3: Presentation of looking time results (Xu and Spelke, 2000).

(4) They did not report whether any of this looking times were any different from the last looking time of the adaptation series – which would indicate that *dishabituation* itself happened at all, or not. Simply comparing different test trials do not indicate dishabituation. Further, having only one occasion to be significant, out of three, could be due to random effects, or due to the detection of an outlier item size, or to an outlier circumference value.

1.1.2. The verbal account

Some studies have found that verbal counting knowledge is indeed inevitable for the formation of the concept of abstract numbers (Mix et al., 1996, Mix et al., 1997, Mix, 1999a and 1999b, Huntley-Fenner and Cannon, 2000, Brannon and Van de Walle, 2001, Rouselle et al., 2004, Rouselle and Noël, 2008).

In the experiment of Mix et al. (1996) 3 and 4-year-old children heard a series of claps. Children were asked to pick a card, out of two cards, which contained as many dots as the number of claps they heard. Numbers 1-4 were used. Verbal counting abilities were measured by the ‘How many’ and by the ‘Give a number’ tasks (Wynn, 1990). In the ‘How many’ task children were asked to count 10 disks and determine the cardinal value. In the ‘Give a number task’ children had to give a certain number of disks from a 15-piece set to the experimenter within number range 1-6. It was found that children could not match the number of items cross-modally before they were able to master the verbal counting system. In a follow-up experiment it was also shown that children could not match superficially dissimilar collections of objects on the basis of their numerosity before they acquired the conventional counting system (Mix 1999b). Overall, researchers concluded that the abstraction of numbers is developing gradually during childhood and is supported by verbal counting knowledge (Mix et al. 1996; Mix 1999a,b).

In contrast to the above, Huntley-Fenner and Cannon (2000) did not find significant relationship between verbal counting and magnitude comparison abilities. In their study 3 to 5-year-olds compared sets of squares ranging from 1 to 15. Scores on the ‘How many’ task did not correlate with comparison performance. It was concluded that number comparison is supported by a preverbal, analogue representation of magnitude which is independent of counting abilities. However, the ‘How high’ task correlated with performance on the comparison task. The authors explained this result by presumed memory effects. They argued that because the ‘How high’ task required children only to recite numbers, it is a measure of memory rather than cardinality knowledge. Hence, children with better memory would not confuse sets with different numbers. The argument was not supported by data in a direct way because memory and its possible relationship to magnitude comparison or to number recitation were not measured.

In another study Brannon and Van de Walle (2001) asked 2 and 3-year-old children to compare the number of foam boxes in two trays. Children had to choose the tray containing more foam boxes than the other one in order to gain stickers as reward. Sets of boxes were called ‘winner’ and ‘looser’ sets, depending on the amount of stickers hidden under each box, and the terms ‘more’ and ‘less’ were never used. Initially, children were trained by doing 1 versus 2 comparisons, and after this they were asked to compare other number pairs with numbers within the range of 1 and 6 (e.g. 1-3, 4-6). Verbal knowledge was measured by the ‘How many’ and by the ‘What is on the card’ tasks. In the ‘How many’ task children had to count and provide a cardinal label for 2, 3, 5 or 6 toys. The dependent measure was the highest number they were able to count correctly. In the ‘What is on the card’ task 2 to 8 apples were shown on a card. The highest number to which children were able to provide the cardinal values was recorded. First, it was found that children without minimal verbal numerical knowledge (those who could not provide at least one correct cardinal response) could not make numerical comparisons. Second, when the subgroup with no cardinal knowledge was removed from the analyses, verbal numerical knowledge did not correlate with magnitude comparison performance any more. Hence, Brannon and Van de Walle (2001) suggested that verbal counting knowledge does contribute to the understanding of the abstract dimension of numbers in children. However, as soon as children capture the idea of discrete numbers with the help of the verbal counting system, their number comparison performance will not rely on verbal numerical abilities any more.

In another thorough study Rousselle et al. (2004) examined the co-development of 3-year-old children’s magnitude comparison skills and counting abilities. Children were asked to compare the number of items in two dot patterns. Two experiments were conducted with conditions controlling for density, contour length and total filled area. 3-year-old children performed above chance level when density or contour length was equated across different sets, but they performed at chance when total filled area was equated across different sets. The authors concluded that at the age of 3 years total filled area or surface is a more salient cue for children than the abstract numerosity of different sets.

In order to examine the relationship of counting and magnitude discrimination abilities, Rousselle et al. (2004) correlated quantity discrimination performance with verbal counting knowledge. They found that children who already mastered and understood the concept of cardinality were better able to compare sets based on their abstract numerosity instead of other perceptual variables. By the examination of individual patterns, it was found that children who performed better on verbal counting tasks, one that necessitates understanding of cardinality in particular, tended to perform above chance on the surface-controlled magnitude comparison task (even if performance on the surface-controlled magnitude comparison task was at chance at the group level).

The authors concluded that the acquisition of verbal counting abilities enables children to understand that numerical quantities are independent of objects' physical properties, like size and luminance. Children who have not yet gone through this conceptual shift would not understand the abstract nature of numbers and would rely on analogue perceptual features in number comparison.

In a recent study, further investigating the relationship of counting and the magnitude representation, Rousselle and Noël (2008) asked 4-, 5- and 6-year-old children to compare two sets of dots or sticks based either on their numerosity or on their overall perceptual size (total filled area). The congruency of the task-relevant and task-irrelevant dimensions was manipulated. For example, in the congruent condition of the numerosity discrimination task collections with larger numerosity occupied larger area than collections with less number of elements. In contrast, in incongruent trials sets with smaller numerosity occupied larger surface area. Results showed that both task-irrelevant perceptual information (in the numerical discrimination task) and task-irrelevant numerical information (in the perceptual discrimination task) influenced children's performance at all ages. Perceptual influence remained stable across ages but numerical influence grew stronger with age, despite the supposedly more mature attentional and inhibition capacities in older children. The divergent developmental pattern of perceptual and numerical processing was interpreted to demonstrate that the non-symbolic numerical feature of the environment is not as salient as non-numerical perceptual features.Contrary to their previous study (Rouselle et al., 2004) no reliable correlations were found between number discrimination performance and counting abilities. It was argued that counting knowledge either contributes to exact calculation only, or that counting knowledge is critical for approximate magnitude processing at the very beginning of its acquisition but its relevance reaches a plateau afterwards.

In summary I can tell that opinions differ in whether verbal counting knowledge contributes to the understanding of magnitude and numerosity relations. On the one hand, Huntley-Fenner and Cannon (2000) assume that magnitude interpretation functions independently of counting. On the other hand, other investigators assume that the abstract dimension of number is discovered and understood only after learning number words (Brannon and Van de Walle, 2001, Mix, 1999a,b). This could happen because labelling quantities by number words emphasizes the abstract similarity between dissimilar objects, for example between three objects and three sounds (Mix et al., 1996). In the absence of this labelling number would not be a salient dimension for preverbal children.

2. The co-development of magnitude discrimination and counting in preschool children

Abstract

4-7-year-old kindergarten children participated in this study investigating the relationship between the development of the magnitude representation, knowledge of number symbols, counting, arithmetic fact retrieval, verbal skills, and numerical and verbal short term memory. The magnitude representation was tested by a non-symbolic numerical Stroop task. Performance on the discrimination task was compared to numerous verbal measures. 4-year-old children, who did not yet possess verbal numerical abilities at a ground level, were unable to discriminate number independently from task-irrelevant perceptual variables. In contrast, 5-7 year old children successfully discriminated number and their performance was compatible with the analogue magnitude representation theory. Number discrimination and verbal skills did not correlate at the level of the whole sample; however, verbal counting knowledge of a group of children with minimal counting knowledge correlated with number discrimination skills when quantities were hard to distinguish (ratio of 2:3). In agreement with the *verbal account* of numerical development, I conclude that verbal knowledge about numbers helps the understanding of the abstract nature of numbers and provides basis for the transition from approximate magnitudes to the recognition of exact and discrete numerosities. Sensitivity to irrelevant features of the task are affected by age and memory, suggesting more general cognitive abilities also play an important part in Stroop-like magnitude comparison.

2.1. Rationale and background

Here my objective was to examine the co-development of the approximate magnitude representation and counting skills in pre-school children. The approximate magnitude representation was examined with a robust non-symbolic magnitude comparison paradigm using an exceptionally large number of trials (64) in each child, with explicitly pitting numerical and non-numerical variables against each other. For the first time, not only accuracy but response times were also recorded of kindergarten children in magnitude comparison. Magnitude comparison performance was compared to counting skills, number knowledge, verbal abilities and short term memory.

The magnitude comparison paradigm employed in the present study is built on the one used by Rouselle and Noël (2008). However, in order to enhance its validity the neutral condition was omitted. This was done because using a neutral condition may result in confounding effects from perceptual variables correlated with number. These variables are interdependent both of each other and of numerosity and it is impossible to control all of them at the same time. For instance, if intensive properties (individual item properties, like item size) are kept equal in a particular trial, extensive properties (properties of the set, like summed surface of all items in a group) will inevitably co-vary with number, and vice versa. Hence, whichever perceptual variable is controlled for in one trial, there will always be another perceptual variable which co-varies with the number of items. The above issue may have affected the results of Rousselle and Noël (2008) in the following way: First, in their neutral condition sets with both larger and smaller numerosity occupied the same amount of area in the numerical task. Second, sets with smaller and larger surface area had the same numerosity in the physical task. Because the outlines (envelope) of the displays were kept equal, density varied as a function of the number of dots and the size of the dots. Hence, the density of the dot displays in the congruent and in the neutral conditions was positively correlated with number. Therefore, density provided a cue positively correlated with number in two-thirds of the trials (in the neutral and in the congruent condition) and it was correlated negatively with number only in the incongruent condition (Figure IV/4).

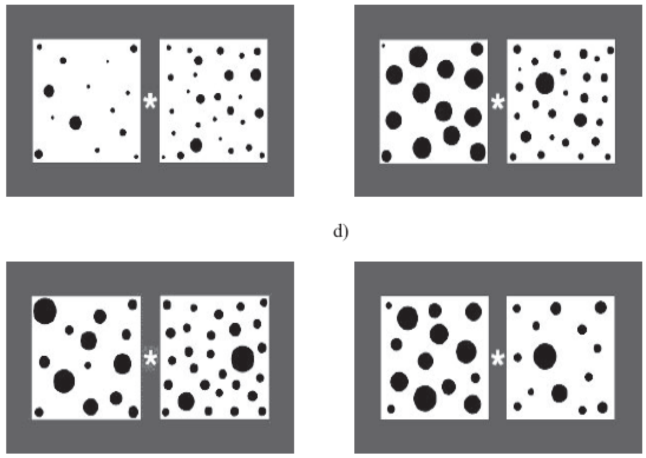


Figure IV/4: The envelopes of the displays were kept equal, density varied as a function of the number of dots and the size of the dots. Hence, the density of the dot displays in the congruent and in the neutral conditions was positively correlated with number (so did so in the 2/3 of the trials). Or, fewer dots were bigger in the neutral and in the incongruent.

This confound may have enhanced the discrimination performance of children who had not been able to focus on abstract numerosity yet. This could explain why the results of Rousselle and Noël (2008) differed from the results of Rouselle et al. (2004).

In order to avoid the above confound here I used only the congruent and incongruent conditions pitting perceptual features against number in exactly 50% of the trials. Congruent and incongruent trials were intermixed with each other. In consequence, attending to a certain perceptual cue will systematically lead to the correct answer in the congruent condition and to the incorrect answer in the incongruent condition. Two perceptual features were controlled for at the same time, so that both perceptual controls were included in the same experiment intermingled in a pseudo-random way within each block and within each subject. Overall surface and circumference were chosen to control for because these were found to be more salient features than number for children (surface: Mix et al., 2002, Feigenson et al., 2002, Rouselle et al., 2004; circumference: Clearfield and Mix, 1999).

A second novelty is that a large number of trials (64) were used and presented all possible trials to each child. This allowed for an ANOVA including Congruency, Ratio, Response side and Type of perceptual control as within-subject factors. Previously either the lack of certain factors or the few number of trials per child impeded to the investigation of interactions between these factors within a single ANOVA design. Most importantly, the effects of ratio and congruency have never been manipulated parametrically in a balanced way and analysed in a single global ANOVA in previous studies of the magnitude comparison task in young children. Specifically, Rouselle and Noël (2008) also included both congruency and ratio in their design, but only congruency and age were used for analysis. Though Halberda and Feigenson (2008) included both factors and age in a non-symbolic number comparison task administered with 3-, 4-, 5-, 6-year-olds and adults, the detailed nature of the inter-relations of age, ratio and congruency were not of the main interest of their study. Barth et al. (2005a) did not analyze ratio and congruency within the same ANOVA either, even though both ratio and congruency manipulation was included in their design and both were shown to exert a significant effect on children's performance. A third novelty is that not only accuracy but also high-precision computer recorded response time (RT) data were also analyzed. To my knowledge, RT has never been analyzed in previous studies with children at this age (e.g. Huntley-Fenner and Cannon (2000) recorded RT with a stopwatch, but results were not reported; Halberda and Feigenson also recorded RT, but results were not reported). In this experiment the high number of trials within each child provided reliable RT data. Information from RT analysis may offer significant advantages to the understanding of cognitive processing in children. RT data sometimes draw a different picture from that of accuracy data – these two measures can be sensitive to different things and serve as a complementary source of information information (in fact, RT was found to be more informative than accuracy in school children's magnitude comparison: De Smedt et al, 2009a, Holloway and Ansari, 2009).

Fourth, not only magnitude discrimination and counting skills were measured, but also used several tests of verbal memory were administered. This was motivated by a growing literature emphasizing the role of memory in the aetiology of numerical disabilities (e.g. Geary, 2004, Passolunghi and Siegel, 2004; Soltész et al. 2007). Rehearsing verbal information, control processes attributed to the central executive (Baddeley and Hitch, 1974) and verbal fluency are all found to be relevant in mathematical disabilities (Geary, 2007; Geary et al, 2007, Hitch and McAuley, 1991, Fürst and Hitch, 2000). Rehearsing verbal information, control processes attributed to the central executive (Baddeley and Hitch, 1974) and verbal fluency are all found to be relevant in mathematical disabilities (Geary et al, 2007, Hitch and McAuley, 1991, Fürst and Hitch, 2000) and in normal numerical development (De Smedt et al, 2009b, Durand et al, 2005).

We can hypothesize three possible scenarios. First, were children relying on perceptual features rather than number per se; they would completely fail to distinguish stimuli on the basis of numerosity in the number discrimination task. In this case children would respond at chance level, or they would systematically always choose the physically larger set of dots instead of choosing the numerically larger set of dots. The second possibility is that children are able to solve the number discrimination task and their performance on this task is correlated with measures of verbal numerical knowledge and counting abilities. The third possible scenario is that knowledge about counting and numbers is not correlated with performance on the number discrimination task because children are able to compare sets based on their numerosity.

2.2. Methods

Subjects. 65 children in a public kindergarten participated in the experiment in, Hungary (Nyíregyháza). Children had working and middle-class background. Children assigned to different age groups entered kindergarten in consecutive years. There were 14 4-year olds (7 boys, mean=4 years, SD=0.14 years), 17 5-year-olds (7 boys, mean=5.6 years, SD=0.28 years), 17 6-year-olds (8 boys, mean=5.96, SD=0.24) and 17 7-year-olds (9 boys, mean=6.88 SD=0.28). Children belonging to different age groups were not separated in the kindergarten most of their time, ie. they played together. In addition, children took part in age-appropriate activities following the kindergarten curriculum in Hungary. Written informed consent was obtained from parents and the study was approved by the institutional ethics committee of the Research Institute for Psychology at the Hungarian Academy of Sciences.

Procedure. Data were collected in two sessions by the children's own kindergarten teacher[17]. During the first session behavioural tests were administered. During the second session the computerized magnitude comparison test was administered.

[17] Here I would like to thank Lívia Szűcs for her most helpful contribution. She organized the testing sessions, motivated and tested all children for this study.

Tests. There were twelve behavioural tests. Tests were grouped into thematic sets. The order of sets was counterbalanced during administration in order to avoid effects arising from general fatigue or boredom. The first thematic set consisted of tests measuring children's *number knowledge*: "Say as many numbers as you can" (number recitation)[18] and a number recognition task (Arabic numerals from 1 to 10 were shown in random order). Children were scored for each number they said and for each Arabic numeral they named correctly. The second set of tests measured *verbal fluency*: "Say as many animals as you can" and "Say as many colours as you can" tasks. Children got a score for every word they listed. The third set of tasks assessed children's *verbal short term memory*: auditory short term memory for numbers and auditory short term memory for words. Children's score was the largest number of items they recited correctly. The fourth set contained tests regarding counting abilities: "Count as far as you can", "How many toys are there" and "Give me a number" tasks. These measures for assessing children's counting abilities followed the design of Wynn (1990, 1992) and Rouselle et al., 2004). Children were scored on the "Count as far as you can" task according to the maximum number they could count to without committing an interchange or an omission. The "How many toys are there" task score reflected the maximum number of elements (toy cars) children could count correctly. The "Give me a number" task score was determined by the maximum number of toys children could correctly select according to the instruction. The number of elements in "How many toys are there" and "Give me a number" was systematically extended until two consecutive errors were made by the child. The fifth set consisted of tasks measuring *verbal problem solving abilities* of children: familiar problems and unfamiliar problems. Familiar problems were additions with the same numbers, e.g. 2+2, 4+4; these additions often come up in kindergarten activities (familiarity was determined by the children's own kindergarten teacher). These additions' results could be simply recited from memory as children have already over-learned them in short rhymes. Contrary, unfamiliar problems could not be easily recited from memory. These problems consisted of similar additions as the familiar problems but were not over-learned by children (e.g. 2+4). Children were

[18] Abbreviations for tests can be found in Table 2. These abbreviations are used in Figure 2 and will not be mentioned later in the text (full test names will be used).

scored by the number of additions they could solve. The difficulty of additions was increased by increasing the problem size (e.g. the simplest problem was 1+1, then 2+2, then 3+3 and so on). The seventh measure characterized children's *understanding of halving*. In this task children were asked to share a salty stick with their peer equally. A drawing of a stick was shown and they were asked to mark the point where they would broke the real stick to share it equally. Performance was measured as millimetre deviation from the middle point.

Magnitude comparison task. Black dots on light yellow background were used as stimuli. Two sets of dots were presented simultaneously on a computer screen (see Figure IV/5).

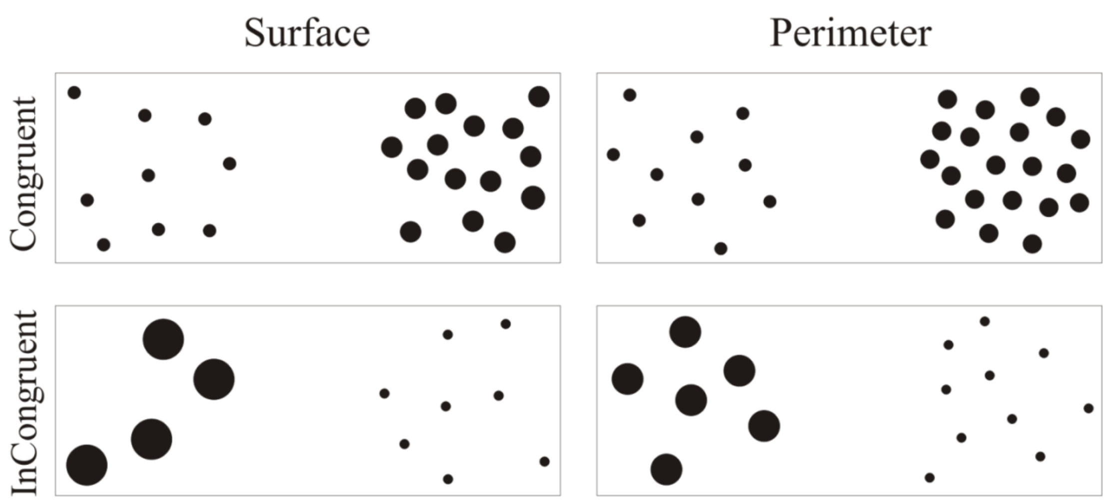


Figure IV/5: sample stimuli.

The sets were separated by 7.5 cm, and were visually easily distinguishable from each other. The overall envelope (corresponding to contour length in Rouselle et al., (2004)) of a set was kept constant at 9×9 cm, as overall envelope has been found to help children even in conditions where overall surface is incongruent with number (Barth et al., 2005a). Children's task was to find out which set contains more dots and press the button at the side of the 'winner' set. Response side was counterbalanced.

The size of dots was constant within set and varied between sets. The individual size of dots and the pattern of dots were randomly varied through pairs of sets. Sets with the same number of items never had the same dot size. Only numerosities above the subitization range were used in order to exclude that object-based attention (or the object file system, Simon, 1997, Kahneman, Treisman and Gibbs, 1992, Uller et al., 1999, Huttenlocher et al., 1994) would be used to complete the comparison task.

The following factors were taken into consideration: (1) The ratio of the number of dots in the two sets. (2) The numerical distance between the number of dot in the two sets. (3) The type of the physical control variable. (4) The congruity of physical control variables and numerosity. (5) The overall numerical sum of items in a display. The ratios and numerical distances for all combinations of numerosities are summarized in Table IV/1. Accordingly, all number pairs used are show in Table 1.

1 : 2	**3 : 5**	**2 : 3**
4 8 *	**6 10 ***	**8 12 ***
6 12 **	**9 15 ****	**12 18 **/°**
10 20 °	**12 20 °**	

Table IV/1: The dot number pairs for each ratio. The numerical distance between dot numbers are the same in dot number pairs marked with * and ** (the numerical distances are 4 and 6, respectively). Ratios are in columns. The overall sum is almost equal for dot number pairs marked with ° (30, 32 and 30)

There was no extreme large numerical distance for the 2:3 ratio because the sum of items would have been much larger than in other conditions (at least 16:24, sum is 40; or 20:30, sum is 50). Rather, I decided to keep the sum of items for the largest numerical distances in the 1:2 and 3:5 ratio conditions to be approximately equal to the sum of items in the distance 6 condition in the 2:3 ratio condition.

Two different physical variables were manipulated as controls: the overall surface (hence, luminance) and the overall circumference (sum of the individual items' circumferences) of the dot groups. These two physical controls were intermixed during stimulus presentation. The ratios of the overall physical sizes (surface in half of the trials and circumference in other half of the trials) of dot sets were congruent or incongruent with the numerical ratio of the dot sets. In the congruent condition the more numerous set was larger in overall physical size than the less numerous set. In the incongruent condition the more numerous set was smaller in overall physical size then the less numerous set. Congruent and incongruent trials were pseudo-randomly intermixed (no more than three of each could follow in a sequence).

In each trial the ratio of perceptual features of the two dot patters was kept the same as their numerical ratio. This was done because if the ratio of the surfaces or circumferences within a set pair were not in accord with the ratio between the numbers in the set pair, the influence of perceptual variables would differ among numerical ratios. For example, if the ratio between the numbers was 1:2 and the ratio between perceptual variables were 2:3, the perceptual difference would be less salient than the numerical difference. This would result in better numerical discrimination performance solely because physical variables would be less distractive. Similarly, if the perceptual ratio were 1:2 and the numerical ratio were 2:3, the perceptual difference would be more salient. Each child was presented with 64 test stimuli pairs, preceded by 8 practice pairs. Trials were separated by funny smiley figures and friendly pictures of animals and soft toys to keep children's attention awake and motivated. Encouraging verbal feedback was given by the kindergarten teacher, independently of performance. Trials were started by the kindergarten teacher when children attended to the screen and agreed to be ready for the next task. All children enjoyed the task and actively looked for participation.

Analysis: behavioural tests. A multiple analysis of variance (MANOVA) was performed on the 12 different tests' raw scores as dependent variables with Age and Gender as independent variables. Results from the twelve tests were entered into univariate F-tests. A Bonferroni type adjustment was performed in order to lower the possibility of the inflated type I error due to multiple comparisons. The critical alpha-level set by the Bonferroni type adjustment was 0.0042 (0.05/12). For the further analyses of significant MANOVA and F-test results post-hoc Scheffé tests were implemented.

Analysis: magnitude comparison task. First, in order to see whether accuracy was significantly different from chance at the group level one-sample t-tests were run against 50%. Second, accuracy and median reaction time (RT) were calculated and entered into ANOVAs with the within subject factors of Side (of the response) × Congruency × Type (of perceptual control) × Ratio and with the between subject factors of Age and Gender.

Correlational analyses. In order to investigate the relationship of counting abilities and magnitude comparison correlational analyses were carried out. Variables were the outcomes of behavioural tests, median RT and accuracy for each condition of the magnitude comparison task (3 levels of Ratio, 2 levels Congruency and 2 levels of Distance). Further, derived variables such as the slope of the Ratio effect (RT and accuracy for smaller ratio minus RT and accuracy for larger ratio), and slope of distance effect (distance effect: RT and accuracy for smaller distance minus RT and accuracy for larger distance) were also used for analysis. Age has been controlled for during the analysis by partial correlation.

Although our sample did not provide enough cases to carry out statistically reliable factor analysis (as optimally at least 300 cases would be required and sample size between 50 and 100 is considered to be very poor, Tabachnik and Fidell, 2007), for the sake of concise representation of correlations a factor analysis was conducted. In fact, factor analysis results were in perfect agreement with correlation results. This is understandable as factor analysis is based on calculations performed on the correlation matrix and extracts factors according to the correlation structure of the data. Hence, the most correlated items will constitute one factor and items that do not correlate with an already extracted factor but do correlate with other items will constitute another factor. That is, items will be ‘grouped’ based on their strength of correlations with other items. A further stepwise regression analysis was conducted in order to investigate the possible predicting value of scores on tests correlating with comparison performance.

Some researchers hypothesize that verbal counting abilities play a significant part in magnitude comparison only in the very beginning of its acquisition. In order to examine this possibility in this data I divided children into “minimal counting knowledge” and age-matched “counting knowledge” groups. These groups were formed by selecting those children who gained 0 score at least in one of the numerical knowledge tasks (“Say as many numbers as you can” and a number recognition task) or in the counting abilities tasks (“Count as far as you can”, “How many toys are there” and “Give me a number”). Age-matched (in months) controls were selected for each child in the counting knowledge group from the pool of remaining children. The performance of the minimal counting knowledge group and of the age-matched control group was contrasted by two-tailed t-tests on behavioural tests and on the magnitude discrimination task. Correlation analyses between test results and magnitude discrimination task measures were conducted separately for both groups.

2.3. Results

2.3.1. Behavioural tests

Normalized test results are plotted in Figure IV/6.

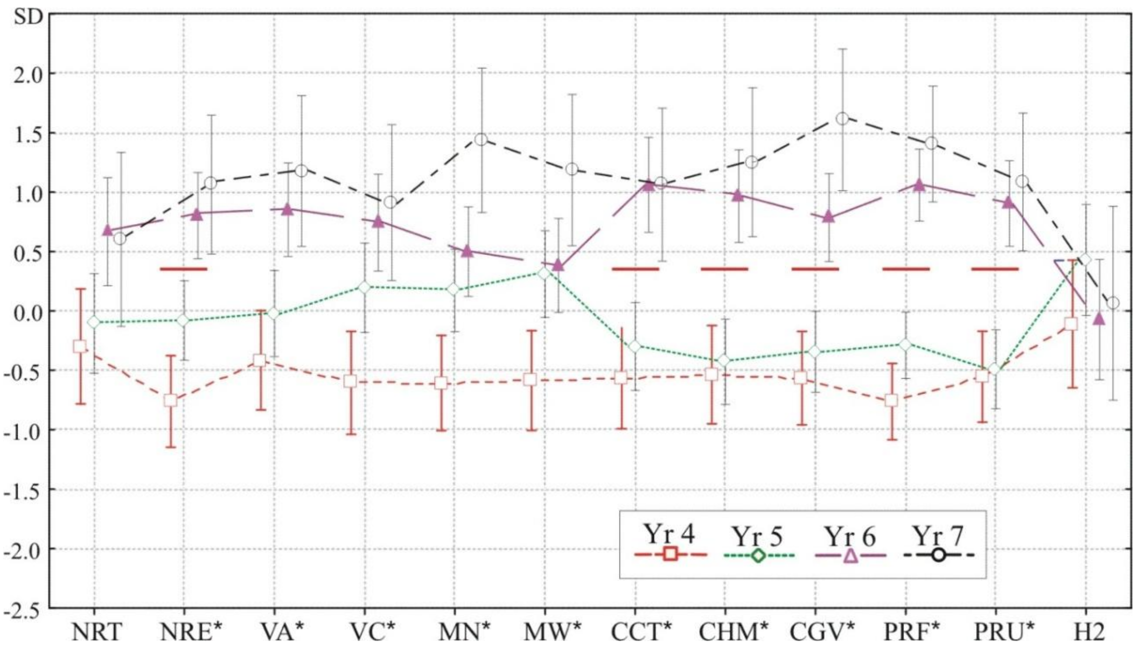


Figure IV/6: Standardized scores (units in SD) for the 12 tests separately for each age group. The critical p value for significance was 0.0042 (see text and Table IV/2). See text and Table IV/2 for abbreviations of tests. Significant age group differences are denoted by red lines. See text for more details.

Data was normalized only to visualize all tests on a common scale independently of the metric differences among different tests. Statistical analyses were conducted on raw test scores. The MANOVA demonstrated that age was a significant factor (Wilks'Λ=0.11, $F(48,175.38)=2.83$, $p<0.0001$). The results of follow-up univariate ANOVAs are given in Table IV/2. The effect of Age was significant in 10 tests. Post-hoc Scheffé tests showed that in 6 out of 10 tests (number recognition, counting abilities and in verbal arithmetic abilities) there was a remarkable developmental breakpoint between the ages of 5 and 6 years: Year 4 and 5 did not differ from each other significantly; neither did year 6 and 7. However, the difference was significant between these two age ranges: both year 6 and year 7 group performed better than any age group in the younger range.

Task	Age		Gender	
	F(4,56)	p	F(1,56)	p
Number knowledge				
- **NRT:** Say as many numbers as you can	3.18	0.0199	0.61	0.4
- ***NRE:** Written (Arabic) number recognition	14.58	***<0.0001***	0.91	0.3
Verbal knowledge				
- ***VA:** Say as many animals as you can	8.66	***<0.0001***	4.7	0.033
- ***VC:** Say as many colours as you can	9.53	***<0.0001***	0.8	0.4
Working memory (verbal)				
- ***MN:** Short term memory for numbers	10.7	***<0.0001***	3.12	0.08
- ***MW:** Short term memory for words	8.19	***<0.0001***	4.78	0.033
Counting abilities				
- ***CCT:** Count as far as you can	12.54	***<0.0001***	1.1	0.3
- ***CHM:** How many objects are there	13.82	***<0.0001***	0.8	0.8
- ***CGV:** Give me a number	15.37	***<0.0001***	0.9	0.3
Verbal counting abilities				
- ***PRF:** Problems – familiar	28.85	***<0.0001***	2.1	0.2
- ***PRU:** Problems – unfamiliar	14.88	***<0.0001***	2.9	0.09
Fractions				
- **H2:** Halving	1.23	0.3	0.04	0.8

Table IV/2: Univariate F-tests for the 12 tests separately. The critical p value for significance was 0.0042 (set by the Bonferroni adjustment). Significant p levels are denoted by bold italic typesetting.

2.3.2. Magnitude comparison

The main effect of **Age** was highly significant on accuracy (Figure IV/7, F(1,58)=11.9, p<0.0001).

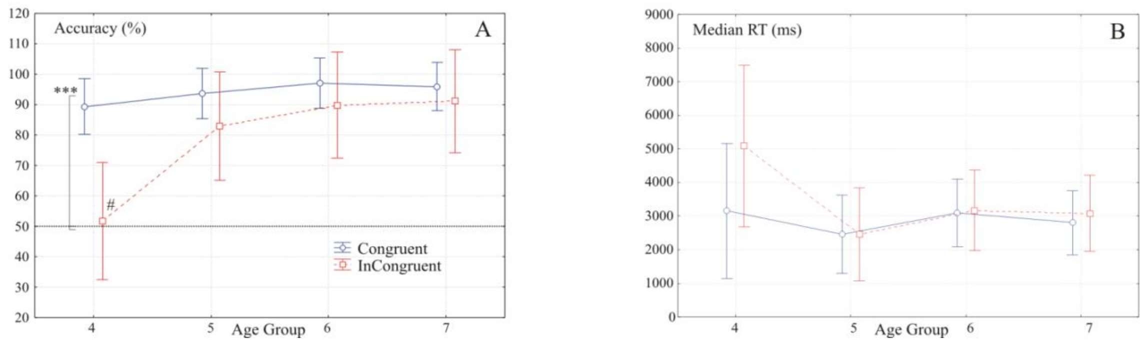


Figure IV/7: A. Congruency effect on accuracy across age groups. Congruency effect reached significance in the 4-years-old group only (p<0.001). One-sample t-tests showed that performance in incongruent condition did not differ from chance in this age group (the non-significant one-sample t-test is denoted by #). B. Significant Age × Congruency interaction (p<0.02).

According to post-hoc Scheffé tests the youngest group omitted significantly more errors than the other three age groups, while there were no significant differences among the three older groups (accuracy and RT data is shown in Table IV/3). The main effect of **Gender** was marginally significant on accuracy (F(1,58)=3.8, p=0.057). Boys committed fewer errors than girls did (89.4% vs. 83.5%). Neither of the above main effects was significant in RT.

Group	Accuracy (%)	Median RT (ms)
4 yrs	70.5	4123
5 yrs	88.3	2465
6 yrs	93.4	3135
7 yrs	93.6	2941

Table IV/3: Summary table for accuracy and reaction time in the magnitude comparison task.

The main effects of **Congruency** (Accuracy: F(1,58)=31.8, p<0.0001; RT: F(1,58)=11, p=0.002) and **Ratio** (Accuracy: F(1,58)=23.4, p<0.0001. RT: F(1,58)=22.5, p<0.0001) were highly significant. Congruent trials were responded faster and with fewer errors than incongruent trials (2880 vs. 3452 ms, 94% vs. 78.9%). More different ratio pairs were responded more accurately and faster than less different ratio pairs (2422, 3062, 4014 ms and 90.8%, 85.5%, 83.2% correct for Ratios 1:2, 3:5 and 2:3, respectively). The **Type** of perceptual control stimulus was significant neither in accuracy nor in RT ($p>0.4$).

The **Age × Congruency** interaction was significant both in accuracy (Figure IV/3A, $F(3,58)=7.5$, $p=0.0002$) and in RT (Figure IV/3B, $F(3,58)=3.96$, $p=0.016$). Pairwise post-hoc comparisons revealed that the effect of Congruency on accuracy decreased with increasing age, reaching statistical significance only in the youngest group ($p=0.0001$) but not in the three older groups. Accordingly, one-sample t-tests showed that all age groups performed significantly above chance in both Congruent and Incongruent conditions except for the 4-year-old group, whose performance did not differ from chance in the Incongruent condition (subsequently for Ratios 1:2, 3:5 and 2:3: 59, 50.5 and 45.5% correct. T-test results in the same order: $t(13)=0.99$, $p>0.3$, $t(13)=0$, $p=1$ and $(13)=-0.42$, $p>0.6$) Post-hoc tests yielded non-significant results for RT data ($p>0.1$).

The **Gender × Congruency** interaction was also significant (Figure IV/8A, Accuracy: $F(3,58)=5.36$; $p=0.024$. Figure IV/8B, RT: $F(3,58)=5.9$; $p=0.02$).

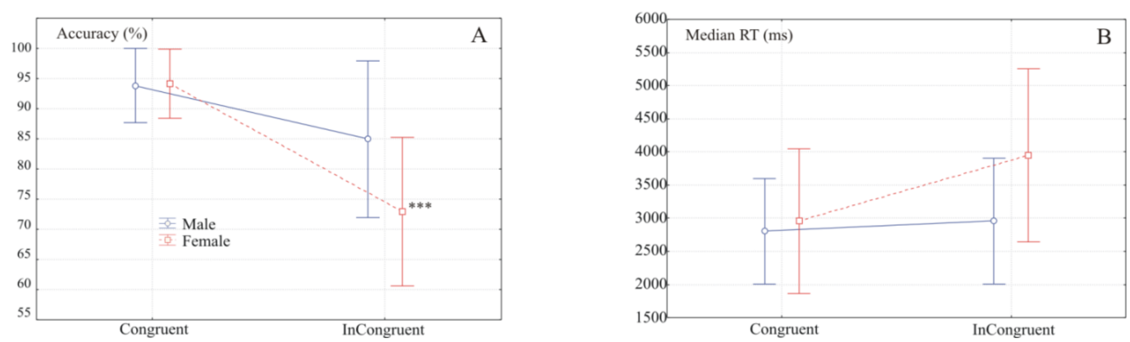
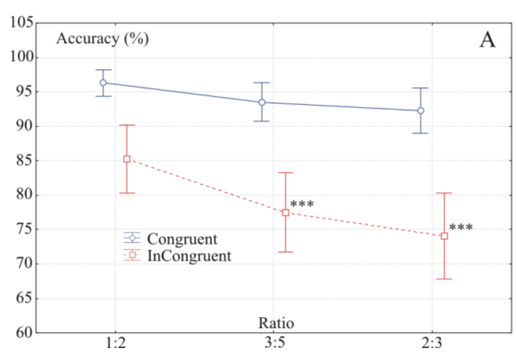


Figure IV/8: A. Congruency × Gender interaction in accuracy data. Females committed more errors incongruent condition than in congruent condition. The same effect was not present in males. B. The same interaction was also significant in RT data ($p<0.05$).

Girls' accuracy was significantly affected by Congruency (94.1% vs. 72.9% in congruent and incongruent trials respectively, $p<0.0002$). In contrast, the Congruency effect was not significant in boys (93.8% and 84.6%, $p>0.16$). No similar effect was found in RT data ($p>0.1$). According to post-hoc comparisons, girls' performance was significantly worse in the incongruent condition than boys' ($p<0.04$). In contrast, the performance of boys and girls did not differ in the congruent condition ($p>0.99$). The three-way interaction of Congruency, Gender and Age was not significant (p>0.3) indicating that the gender difference in the congruency effect is stable across age groups. The possibility of a gender-driven chance-performance in the 4-year-old group was tested by one-sample t-tests comparing boys' and girls' accuracy separately to 50%. Both girls and boys performed at chance in the incongruent condition, in all three ratios in the 4-year-old group (p>0.19). And both girls and boys performed above chance in the congruent condition (p<0.01) in the 4-year-old group. Both boys and girls performed above chance in both congruent and incongruent conditions in the older age groups (p<0.01). However, post-hoc pairwise comparisons (Tukey-Kramer test) revealed that 4-year-old girls' accuracy performance is significantly worse than all the other data points (p<0.001, from both girls' and boys' performance in both conditions across all age groups), and considerably different (although not yet significantly) from 4-year-old boys' performance in the incongruent condition (4-year-old girls: 36.9%, 4-year-old boys: 66.5%, p<0.07). No such result emerged from the RT data.

The **Congruency × Ratio** interaction was significant in accuracy (Figure IV/9A): F(2,116)=4.7; p=0.01) and it was marginally significant in RT (Figure IV/9B, p=0.066).



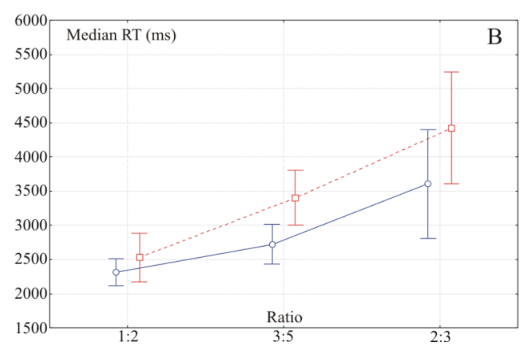


Figure IV/9: A. Congruency × Ratio interaction in accuracy. Ratio effect was significant only in the Incongruent condition (*** denotes p<0.001 significance level). B. The same interaction was marginally significant in RT data (p=0.066).

Post-hoc tests revealed that there was a significant Ratio effect in the accuracy data only in the incongruent trials (p<0.0005 between ratio 1:2 and ratio 3:5; and p<0.0002 between ratio 1:2 and ratio 2:3) but not in the congruent trials. In RT, planned comparisons revealed that the ratio effect was significant in both congruent and incongruent conditions ($p<0.001$ between ratio 1:2 and 3:5; and between ratio 2:3 and 3:5).

Interestingly, the **Age × Gender** interaction was significant in RT (Figure IV/10, $F(3,35)=3.3$, $p=0.029$) but not in accuracy.

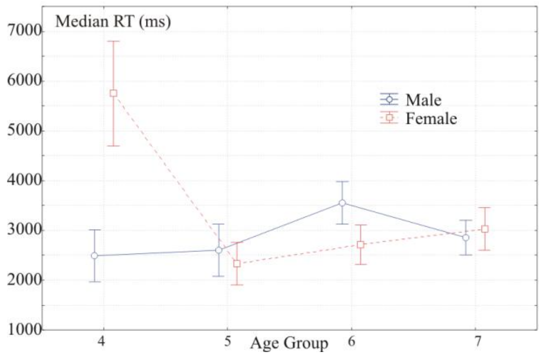


Figure IV/10: Age × Gender interaction in RT ($p<0.05$).

Boys responded faster than girls in the 4-year-old group (2495 vs. 5750 ms, the difference is 3254 ms). This discrepancy than disappeared in the older groups (2605 vs. 2325 ms, 3558 vs. 2712 ms, and 2859 vs. 3023 ms, respectively). None of the post-hoc comparisons was significant.

2.3.3. Correlations

The matrix of correlation coefficients (age was controlled) for test results is shown in Table IV/4. The correlation coefficient matrix for the magnitude task results is shown in Table IV/5. Only the two short term memory measures showed correlations with some error measures of the magnitude comparison task (Table IV/6). Strikingly, there were no any other correlations between behavioural tests and the magnitude comparison task.

	NRT	NRE	VA	VC	MN	MW	CCT	CHM	CGV	PR1	PR2	H2
NRT	1.00	0.27	0.20	0.24	0.22	0.16	0.37	0.31	0.21	0.13	0.33	-0.04
NRE	**0.27**	1.00	0.10	0.35	0.19	0.28	0.25	0.26	0.17	0.14	0.07	-0.06
VA	0.20	0.10	1.00	0.39	0.24	0.21	0.33	0.34	0.29	0.40	0.48	-0.05
VC	0.24	**0.35**	**0.39**	1.00	0.12	0.16	0.08	-0.01	0.00	0.20	0.05	0.04
MN	0.22	0.19	0.24	0.12	1.00	0.46	0.12	0.21	0.18	0.11	0.15	-0.09
MW	0.16	**0.28**	0.21	0.16	**0.46**	1.00	-0.08	0.00	0.04	-0.02	0.10	0.00
CCT	**0.37**	0.25	**0.33**	0.08	0.12	-0.08	1.00	0.82	0.77	0.45	0.46	-0.13
CHM	**0.31**	**0.26**	**0.34**	-0.01	0.21	0.00	**0.82**	1.00	0.84	0.37	0.56	-0.16
CGV	0.21	0.17	**0.29**	0.00	0.18	0.04	**0.77**	**0.84**	1.00	0.28	0.47	-0.14
PRF	0.13	0.14	**0.40**	0.20	0.11	-0.02	**0.45**	**0.37**	**0.28**	1.00	0.51	-0.29
PRU	**0.33**	0.07	**0.48**	0.05	0.15	0.10	**0.46**	**0.56**	**0.47**	**0.51**	1.00	-0.19
H2	-0.04	-0.06	-0.05	0.04	-0.09	0.00	-0.13	-0.16	-0.14	**-0.29**	-0.19	1.00

Table IV/4: R values from partial correlations for tests. Bold typesetting indicates significant (p<0.05) correlations. For further details see text.

	RT-CON	RT-INCON	RT-RAT1	RT-RAT2	RT-RAT3	RT-ALL	ERR-CON	ERR-INCON	ERR-RAT1	ERR-RAT2	ERR-RAT3
RT_CON	1.00	0.54	0.71	0.83	0.85	0.92	-0.26	-0.14	-0.18	-0.17	-0.24
RT_NCON	**0.54**	1.00	0.68	0.68	0.63	0.74	0.00	-0.27	-0.21	-0.22	-0.22
RT_RAT1	**0.71**	**0.68**	1.00	0.68	0.63	0.80	-0.33	-0.32	-0.27	-0.38	-0.42
RT_RAT2	**0.83**	**0.68**	**0.68**	1.00	0.68	0.88	-0.32	-0.26	-0.35	-0.28	-0.32
RT_RAT3	**0.85**	**0.63**	**0.63**	**0.68**	1.00	0.93	0.00	-0.06	0.02	-0.04	-0.10
RT_ALL	**0.92**	**0.74**	**0.80**	**0.88**	**0.93**	1.00	-0.18	-0.19	-0.18	-0.20	-0.25
ERR_CON	**-0.26**	0.00	**-0.33**	**-0.32**	0.00	-0.18	1.00	0.25	0.54	0.52	0.57
ERR_INCON	-0.14	**-0.27**	**-0.32**	**-0.26**	-0.06	-0.19	**0.25**	1.00	0.85	0.89	0.86
ERR-RAT1	-0.18	-0.21	**-0.27**	**-0.35**	0.02	-0.18	**0.54**	**0.85**	1.00	0.81	0.76
ERR-RAT2	-0.17	-0.22	**-0.38**	**-0.28**	-0.04	-0.20	**0.52**	**0.89**	**0.81**	1.00	0.81
ERR-RAT3	-0.24	-0.22	**-0.42**	**-0.32**	-0.10	-0.25	**0.57**	**0.86**	**0.76**	**0.81**	1.00
ERR-ALL	-0.21	-0.23	**-0.39**	**-0.34**	-0.05	-0.22	**0.58**	**0.93**	**0.91**	**0.94**	**0.93**

Table IV/5: R values from partial correlations for behavioural measurements. Bold typesetting indicates significant (p<0.05) correlations

	MN	MW
ERR_INCON	0.22	**0.28**
ERR_RAT1	**0.27**	**0.32**
ERR_RAT2	0.21	0.24
ERR_RAT3	**0.25**	**0.28**
ERR_ALL	**0.26**	**0.30**

Table IV/6: R values from partial correlations among tests and magnitude task measurements. Bold typesetting indicates significant ($p<0.05$) correlations.

	1	2	3
NRT	0.239429	0.539796	0.117110
NRE	0.190808	**0.705857**	0.145507
VA	0.050910	**0.712302**	0.258825
VC	0.120982	0.557922	0.268724
MN	0.175689	0.551612	0.379478
MW	0.170932	0.405931	0.409953
CCT	0.049560	**0.909566**	0.164922
CHM	0.036905	**0.889526**	0.213583
CGV	0.044615	**0.867434**	0.223804
PR1	0.058252	**0.824896**	0.223152
PR2	0.143934	**0.807914**	0.217954
H2	-0.022292	-0.101818	-0.107875
CONGR	**-0.930375**	-0.093792	-0.103606
INCON	**-0.722560**	-0.013829	-0.208961
RAT1	**-0.724875**	-0.188580	-0.400200
RAT2	**-0.855544**	-0.115420	-0.317704
RAT3	**-0.924577**	0.053293	0.126029
ALLZU	**-0.982134**	-0.056095	-0.133725
ERR_CONGR	0.173563	0.117307	0.640845
ERR_INCON	0.131973	0.220069	**0.892832**
ERR_RAT1	0.128044	0.162317	**0.909502**
ERR_RAT2	0.143302	0.232862	**0.911971**
ERR_RAT3	0.195263	0.224094	**0.892648**
ERR_ALL	0.167608	0.221621	**0.954582**
CONd	0.021703	-0.043851	0.007584
ERR_CONd	-0.174141	0.079530	-0.021073
RATd	0.032301	-0.039026	0.079926
ERR_RATd	-0.141396	-0.125170	0.014883
DISTd	-0.499843	0.069788	-0.178550
ERR_DISTd	0.064107	0.173468	-0.107101
Expl. Var.	9.784019	6.247673	5.951855

Table IV/7: Factor loadings. Extraction method: principal components were extracted and varimax rotation was applied. Marked loadings are > 0.7.

The factor loadings from the additional factor analysis are shown in Table IV/7. RT measures (congruence and incongruence effects, levels of ratio) were highly intercorrelated and formed one factor. A second factor was formed by *number recognition, verbal fluency - animals*, counting abilities: *"Count as far as you can", "How many toys are there"* and *"Give me a number"*, and *familiar problems* and *unfamiliar problems*. The third factor included the error measures of the magnitude task.

It is to note that a clear trade-off between RT and accuracy can be seen in Table IV/5. RTs and accuracy are in a strong negative relationship. This means that the faster the response the lower the number of correct answers.

2.3.4. Regressions

Stepwise regression analysis was carried out entering the overall error rate of magnitude task as dependent variable (DV) and *short term memory for numbers* and *short term memory for words* as independent (predictor) variables (IVs). These two tasks were chosen because these are the only tests which correlated with some of the magnitude comparison task measures (namely with some of the error related measures). Regression analysis is the most effective if selected IVs are correlated with the DV (Tabachnick and Fidel, 2007). *Short term memory for words* accounted for the 19% of the variance in DV (adjusted $R^2 = 0.19$; $F(1,64)=16.4$, $p<0.001$, $SE=2.42$). *Short term memory for numbers* did not reach the level of significance in explaining the remainder variance in order to be taken into the regression model.

2.3.5. Comparison of groups with minimal counting knowledge and age-matched controls

13 children were selected upon the criteria that at least one their number knowledge or counting abilities score was zero (minimal counting knowledge group). An age-matched group of 13 children was selected randomly from the remaining children (control group). The age range was 46-73 months including five 4-years-olds, five 5-years-olds and three 6-years-olds in both groups. The mean ages of the groups were 58.76 and 58.69 months (minimal counting knowledge and control group, respectively). The two groups did not differ from each other in age (p>0.98). Gender was balanced within and between groups (minimal counting knowledge group: 3 boys

and 2 girls of age 4; 4 boys, 1 girl in age 5; 2 boys, 1 girl in age 6. Control group: 2 boys, 3 girls in age 4; 3 boys, 2 girls in age 5; 2 boys, 1 girl in age 6).

Groups did not differ in any measures of the magnitude comparison task. In other tests control children performed better in the *number recognition task* (0.8 and 4.7 mean scores for the minimal counting knowledge group and for the control group respectively; $t(24)=-3.12$, $p<0.005$). Controls were also better in the *„say as many animals as you can"* (verbal abilities) tasks (means: 7.2 and 11.5, $t(24)=-2.65$, $p<0.02$) and marginally better in the *„count as far as you can"* (mean: 14.7 and 28.2, $t(24)=-1.95$, $p=0.06$). There were no other differences in any other tests.

Accuracy in the magnitude comparison task was tested against chance. Accuracy of children in the minimal counting knowledge group only marginally (68.3%, $p=0.06$) differed from chance in the Incongruent condition when the ratio was the largest (2:3 ratio). Similarly, the accuracy of the age-matched control group was only slightly significantly above chance in the above condition (71.2%, $p=0.04$), and did not differ from the minimal counting knowledge group's accuracy (according to the group comparison t-tests). Accuracy in all other conditions was significantly different from chance in both groups.

A partial correlation analysis was also performed on both groups' data. There were no correlations between behavioural tests and magnitude discrimination performance in the control group. In contrast, in the minimal counting knowledge group there were such correlations: First, RT for the harder ratios negatively correlated with the highest number („*Say as many numbers as you can*") the child could recite (3:5 ratio: $r^2 = -0.58$. 2:3 ratio: -0.64; $p<0.05$ for both). The more numbers a child could recite, the faster (s)he was to choose the larger numerosity in the most difficult ratio conditions. Furthermore, the more numbers a child could recite, the smaller was the effect of ratio on RT (measured by the slope of the ratio effect: $r^2 = 0.61$ $p<0.05$).

3.4. Discussion of the results

3.4.1. Congruency effects in magnitude comparison

Congruency effects were of interest because they can attest whether children attended to numerosity or to irrelevant perceptual variables. A significant congruency effect was found which was driven by 4-year-old girls. Closer inspection revealed that both 4-year-old girls and boys performed at chance level in the incongruent condition of the magnitude discrimination task, but girls showed worse performance than boys. In contrast, older children performed above chance. Considering these results I conclude that children at 4 years of age are inevitably influenced by the task-irrelevant continuous perceptual properties of the display and they are not able to perform intentional numerical judgments independently of the items' physical appearance. Our data demonstrates that children with minimal counting knowledge do not differ from age-matched controls in the effect of congruency. Individual inspection revealed that both 4-years-olds with minimal counting knowledge and 4-years-olds with counting knowledge showed the congruency effect. That is, counting knowledge did not affect the congruency effect. This suggests that while 4-year-olds may already understand counting principles, such as one-to-one correspondence and cardinality, task-irrelevant continuous perceptual variables influence their performance in numerical comparison.

The inability of 4-year-olds to avoid the effect of perceptual variables apparently contradicts findings according to which even infants are able to discriminate the numerosity of dot patters when perceptual variables were controlled for (e.g. Xu and Spelke, 2000; Feigenson et al. 2002a; Xu et al., 2005). There may be a number of explanations for this discrepancy. First, experiments with infants used the habituation paradigm because infants cannot be explicitly instructed to attend to numerosity. In contrast, experiments with older children instructed children to attend to number. Therefore, it is a possibility that both infants and 4-year-old children are inherently sensitive to number but they are not yet able to access it at an intentional conscious level (Rouselle and Noël, 2008, Karmiloff-Smith, 1997; Mix, 1999). Similarly, it is also possible that young children cannot consciously direct attention to number, yet.

Differences between tasks also have to be taken into consideration when explaining why 4-year-old children's number discrimination performance may be worse than that of infants. In particular, certain types of tasks or experimental designs may emphasize numerosity more than others, guiding children's attention toward this dimension (for instance van Loosbroek and Smitsman, 1990 used moving, and Wynn et al., 2002 used moving and grouped stimuli; or Barth et al., 2005a and Rouselle and Noël, 2008, using incongruent condition). For example, Xu and Spelke (2000) kept only numerosity constant and varied other variables during habituation. Hence, changing everything else but number may have emphasized number as the relevant dimension for young children in Xu and Spelke's experiment (Brannon et al., 2004). In contrast, in other designs the relevant numerical dimension remains undefined and even if children are sensitive to numerical properties, perceptual variables may be more salient and more attractive for them. For example, Clearfield and Mix (1999) kept both numerosity and perceptual variables constant during habituation, which probably prevented children from explicitly directing attention to number and therefore they may not have realized that it is the number of the stimuli what counts.

3.4.2. Congruency effects: development

There was a clear sign of developmental progression in number discrimination between the ages of 4 and 5. While 4-year-old children performed at chance level in the incongruent condition, children older than 4 years of age performed significantly above chance in the incongruent condition. This change was independent of and preceded the developmental change in verbal and counting abilities, which happened between ages 5 and 6. The lack of correlation between congruency effects and counting abilities (see Table IV/6), suggests that children's sensitivity to incongruent irrelevant stimulus dimensions was independent from number knowledge. Factor analysis also demonstrated that congruency effects and counting abilities loaded to two different factors. This progression can potentially be explained by both numerical and non-numerical factors. One possibility is that the processing of relevant numerical information became more efficient in older children. This may have also increased the saliency of numerical information. An alternative, non-numerical, explanation can be that older children had more mature cognitive and motor inhibition capacities than younger children and this contributed to the more efficient processing of task-relevant information in general.

There are arguments for both explanations. First, there is ample evidence that older children perform better in a range of Stroop-like tasks than younger children (Gerstadt, Hong and Diamond, 1994; Prevor and Diamond, 2005; Halberda and Feigenson, 2008), and that adult level performance is not reached on Stroop tasks till the age of 21 (Huizinga et al. 2006). Moreover, it has been shown that automatic processing of irrelevant physical properties is inevitable even in adults and these properties exert a significant effect even on adult's performance in dot group comparison (Hurewitz et al., 2006). Furthermore, have shown that 9-year-old primary school children demonstrate substantial incorrect motor activation in the incongruent conditions of symbolic and non-symbolic Stroop tasks (Szűcs et al. 2007, 2008). The above evidence suggest that general conflict resolution skills, most probably cognitive and motor inhibition abilities substantially contribute to performance in Stroop-like tasks where stimuli have both task-relevant and task-irrelevant properties. This also raises the possibility that the way of recording discrimination performance can also contribute to the apparent discrepancy between infant and 4-year-old performance. Infant experiments use habituation techniques and measure looking times. In contrast, experiments with older children require manual responses. Hence, motor inhibition abilities related to effector control can significantly affect results with older children whereas they do not play a role in infant studies.

Second, Rouselle and Noël (2008) tested performance in both numerical and physical size comparison. They found that numerical information exerted larger influence on older children's comparison even if numerical information was task-irrelevant (in the physical size comparison task). This suggests that the automatic processing of numerical information developed with age. Importantly, Rouselle and Noël (2008) measured general inhibition capacities with the Day-Night Stroop task where children have to say 'day' when they see the picture of the Moon and say 'Night' when they see the picture of the Sun (Gerstadt et al., 1994). In agreement with previous child Stroop results it was found that 4-year-olds were less accurate and slower than their 5- and 6-year-old peers. However, the developmental effect found in the Day-

Night task did not correlate with the interference effect on the magnitude discrimination task. This suggests that developing inhibition abilities may not be able to fully explain decreasing interference effects in number discrimination. I conclude that probably both growing inhibition capacities and more efficient automatic number processing mechanisms contribute to the decreasing influence of irrelevant perceptual variables in older children.

3.4.3. Ratio and congruency effects in magnitude comparison

The Ratio effect on accuracy and reaction time was highly significant in all age groups. This is consistent with previous and recent findings (Huntley-Fenner and Cannon, 2000, Rouselle et al., 2004, Barth et al., 2005a,b, Halberda and Feigenson, 2008).

Under tightly controlled circumstances it has been found that 4-year-olds are still unable to disregard irrelevant perceptual features in non-symbolic magnitude discrimination. Their performance was at chance when physical size was in conflict with numerical information (see Figure IV/3). A recent study also found significant interactions among age, congruency and ratio (Halberda and Feigenson, 2008). However, the interplay of age, congruency and ratio was not analysed in detail. For instance, when comparing performance against chance in age groups, the authors collapsed across all the ratios; however, a detailed inspection of Figure 2. (p. 1461, Halberda and Feigenson, 2008) reveals that young children were at chance with more difficult ratios, especially in incongruent condition. Though not reporting detailed results, the authors account for the chance performance by younger children arguing that these children simply 'ceased' to rely on their ANS in the more difficult situations involving larger ratios. Whether the incongruence of perceptual information played a significant role in drawing children's performance to chance level is not reported.

We found that 4-year-olds were at chance, but only in the incongruent and not in the congruent condition. Further, and most importantly, it has been found that 4-year-olds performance in this condition was modulated by verbal numerical knowledge (see the next section).

3.4.4. Verbal counting abilities and magnitude comparison

The most important question was how magnitude discrimination co-develops with counting abilities. Looking at the whole sample, there were no correlation between performance on magnitude discrimination and children's counting abilities and number knowledge. Similarly to us, Huntley-Fenner and Cannon (2000) also found that counting knowledge (when children have to apply the cardinal value after a counting sequence) did not correlate with magnitude discrimination, but number recitation (saying number words in order), predicted performance in the magnitude discrimination task. Huntley-Fenner and Cannon (2000) concluded that the correlation between number recitation and magnitude discrimination was due to memory capabilities instead of verbal counting knowledge, because number recitation and counting did not correlate with each other. They also argued that memory effects could also explain why some children know more number words and are less prone to confuse numerical sets than others. A weakness of the above explanation was that memory was not measured explicitly by Huntley-Fenner and Cannon (2000). In contrast to the above, here not only counting and number knowledge but also memory span was measured. No correlations were found between memory and number knowledge in our sample while number knowledge correlated with the ratio effect and with counting knowledge. Memory correlated with the number of errors in the magnitude comparison task. Therefore, our data suggest that number knowledge did not reflect memory capacity as suggested by Huntley-Fenner and Cannon (2000). Rather, on the one hand, I conclude that number knowledge is a measure independent from memory. On the other hand, correlation between memory and the number of errors may reflect more general cognitive abilities responsible for control processes required when stimulus dimensions are in conflict.

Further, the developmental change in counting and verbal problem solving abilities between 5 and 6 years was not accompanied by any changes in magnitude comparison performance. This sudden change in verbal numerical abilities could be clearly due to educational effects as children of that age were participating in planned preschool activities in the kindergarten, emphasizing basic mathematical training in particular. Further, these counting and verbal problem solving abilities contributed to the same factor (see Table IV/7) together with verbal fluency. Hence, it seems reasonable to assume that that verbal ability was a common factor behind the development of the above measures.

In a secondary analysis of 13 children with minimal counting knowledge I found that a number knowledge measure (recitation of number words) correlated with the ratio effect in the magnitude discrimination task. It is important to note that the effect of number knowledge was present only to RT and not in accuracy, suggesting that the two measures (RT and accuracy) are sensitive to different aspects of cognitive development. No similar correlations were found with other counting measures.

Furthermore, children with minimal counting knowledge could not perform above chance in the hardest ratio (2:3) in the incongruent condition, while the age-matched control group performed above chance in all ratio and congruency conditions (However, the between-group performance difference did not reach significance [p>0.9] and was much more robust among age groups). The above suggest that there may be a relationship between number knowledge and number discrimination skills in children with counting knowledge. This conclusion is in agreement with previous investigators according to whom there is correlation between language-based number skills and magnitude comparison performance at the early stage of the acquisition of verbal number knowledge (Rouselle et al., 2004, Brannon and Van de Walle, 2001, Mix et al., 1996, Mix, 1999a, Mix 1999b, Siegel, 1977).

One possibility is that refined number discrimination is related to knowledge about the abstract dimension of number. For example, better (abstract) number knowledge can help children to understand numerical differences more effectively, especially when numerical values are close to each other. It is also possible that better number knowledge can help in drawing children's attention to number more effectively (Brannon and Van de Walle, 2001). A particular theoretical approach suggests that the magnitude representation which supports numerical comparisons is not specific to numerical stimuli; rather it is a broad cognitive domain which applies to most quantitative dimensions (for a review, see Cantlon et al., in press). This broad cognitive domain activates the same computational mechanisms to analogue physical properties like length, luminance or approximate number. For instance, strong interference of conflicting physical information and approximate number information may result from the automatic and involuntary use of the same computational mechanism on both dimensions. Among natural circumstances, this common or mainly overlapping approximation mechanism may not hamper the success of the individual because number and size strongly tend to correlate with each other. Identification and attention towards one specified dimension, like number, while the otherwise naturally correlated property, like size is in strong contradiction, may require additional control processes. I suggest that number knowledge helps children to identify and to separate numerical information voluntarily; they are on the way from approximate magnitudes towards the recognition of exact numbers.

The acquisition of counting and the retrieval of number facts are considered very important objectives of early schooling, and higher level mathematics skills are built on exact numbers (Feigenson et al. 2004). Therefore, it is highly relevant that several developmental studies were unable to show any relationship between the development of magnitude discrimination (thought to index the magnitude representation) and verbal number skills after the acquisition of minimal verbal number knowledge at around age 4.

The above suggest that the development of magnitude discrimination and counting knowledge correlate with each other only at the very early stages of arithmetical enculturation. This implies that if magnitude representation is really a basic and essential arithmetic skill than it reaches its functional maturity very early, at around age four. Following this train of thought, the development and further maturation (Sekuler and Mierkiewitz, 1977) of the magnitude discrimination may not play a serious role in the acquisition of school mathematics. Also, in some recent studies symbolic magnitude comparison was found to be related to arithmetic performance while non-symbolic magnitude comparison was not (Holloway and Ansari, 2009, Rouselle and Noël, 2007), suggesting that the efficiency of the associational link between exact number symbols and the meaning of this symbols is more important in school arithmetic than non-symbolic magnitude representation.

Contrary to the previous assumption, a recent paper asserts that individual differences in the acuity of numerical magnitude representation predict mathematics performance during school years (Halberda et al., 2008). The Halberda et al. (2008)

study is the first attempt to fill the gap in the literature and explicitly measure the relation of number representation and later maths performance. The authors measured the acuity of the Approximate Number System (ANS, Halberda et al., 2008) of 64 children at age 14 and correlated these results retrospectively with their earlier maths performance from 5 to 11 years. ANS acuity was tested by a non-symbolic number comparison task. Further, they administered IQ tests and rapid colour naming tests in order to control for general abilities – intelligence and task demands – in their correlational analyses. They found that ANS acuity measured at 14 years of age correlated with earlier maths performance each year from 5 to 11. ANS acuity at 14 years of age also correlated with symbolic maths performance measured at 8 years of age when IQ and task demands were controlled for. Based on these correlations, the authors assume that "...the ANS may have a causal role in determining individual maths achievement." (p. 667). However, there are some methodological difficulties the study might not have fully overcome yet. First, a statistical correlation never indicates causal relation among variables, let alone the direction of causation. Second, ANS acuity was measured only at age 14 while maths performance was measured at earlier time points (5-11 years). So forth, the reversed causal relation, namely that maths performance in early school years predicts ANS acuity during early adolescence, would also be a feasible explanation of the results. This latter assumption can also be supported by studies of populations who speak innumerate languages and do not receive any formal education. Performance of Amazonian native groups on number comparison tasks suggest the functioning of the analogue magnitude representation (Pica et al., 2004; Gordon, 2004). However, their ANS acuity is weaker than their educated Western peers', showing that education and the Western numerical culture (e.g. use of exact number words above 3) cause differences in ANS acuity (Pica et al., 2004). Third, general abilities were controlled for only at age 8 and not at other ages. Furthermore, a closer look at the results reveals that intelligence correlated ($p<0.04$) with ANS acuity when symbolic math achievement was controlled for (see Table 2 on p. 667) and visual working memory also (though marginally: $p=0.06$) correlated with ANS acuity. These additional results could also suggest that possibly more than one and also more general abilities are involved in symbolic math besides ANS acuity. The possibility of more complex causal relations among ANS, symbolic maths and general abilities cannot be ignored either. For example, from the Halberda et al. (2008) study the reader finds no information about whether IQ and symbolic math were in any relationship – the authors reported the relation between ANS acuity and IQ only. If ANS acuity is hypothesized to be a predictor of symbolic maths achievement, it is of great interest whether more general abilities, like IQ, would or would not explain more variance of symbolic maths achievement than ANS acuity does.

On the other hand, it may be possible that grave deficits of fundamental magnitude discrimination skills may have a devastating effect on mathematics knowledge resulting in a certain type of dyscalculia (Butterworth 1999). However, it also seems that deficits of the magnitude representation are not necessary prerequisites of all kinds of dyscalculia because children and adolescents with dyscalculia can have perfectly normal behavioural number discrimination skills (Soltész et al. 2007; Kucian et al. 2006; Price et al. 2007, Rubinstein and Henik, 2005, 2006). Uncertainties suggest that further research should specify the exact relationship of magnitude discrimination and higher-level mathematics skills required for school mathematics performance during the life-span.

4. Conclusion: Magnitude representation, language, numbers, and development

In conclusion, number discrimination shows reliable ratio and distance effects in 4-7 year-old children and number comparison is supported by the analogue magnitude representation. The carefully controlled perceptual dimensions of the present study assured that children made their decisions upon numerical value and not upon summed physical properties of the stimuli. Although, 4-year-olds were not able perform above chance when perceptual features were in conflict with numerosity. I suppose that this stems from the development of more general cognitive abilities responsible for response execution and inhibition processes, which are required in tasks with such conflicting stimuli dimensions. Tasks at this level of complexity (even if there are more than one stimulus property, see for example Huttenlocher et al., 1994) can be very demanding for young children and tax some more domain-general abilities besides numerical perception per se. These domain-general abilities should be taken into consideration in any further field-specific developmental research.

The development of verbal counting and number knowledge correlates with magnitude comparison only at the first stages of the acquisition of verbal counting and number knowledge. We can conclude that language abilities in fact contribute to the understanding of the abstract nature of numbers, but only at the first steps towards its understanding (Brannon and Van de Walle, 2001). Explicit numerical enculturation, a peculiarity of the human race, helps us both to refine our numerical approximation procedures and later to implement exact and symbolic calculations.

V. Interference between physical and numerical magnitudes: A developmental study[19]

Abstract

While there are large differences between children and adults in the response time in number comparison tasks, a seminal study by Temple and Posner (1998) has shown that numerical information is extracted in children just as fast as in adults; it is rather during the response organization/execution phase when children lag behind adults. Furthermore, attention and inhibition plays also a considerable role in children's performance in numerical tasks (Szűcs, Soltész et al., 2007). The present study investigates the speed of numerical processing in grade 3-5 children and in adults, by measuring the timing of the numerical distance effect in ERPs in a numerical Stroop paradigm. Besides the speed of number processing, the role of the inhibition of irrelevant features is also examined. According to the results, children extract numerical information almost as fast as adults. Furthermore, children are more susceptible to interfering information than adults, which has to be considered in any numerical tasks with children.

1. Rationale and background

1.1. Speed of number processing: the numerical distance effect

The most commonly used marker of the analogue magnitude representation is the *numerical distance effect* (see chapter I**,** 1.1.). In number comparison tasks, where for example one has to indicate whether the presented number is smaller or larger than five, was found that the numerical distance effect in children is very similar to that of adults. This suggests that numerical magnitudes are represented in a phenomenologically similar way in children and in adults (Sekuler and Mierkiewitz, 1977; Temple and

[19] This study is submitted (to Learning and Individual Differences)

Posner, 1998). However, children are much slower in magnitude comparison tasks, which could lead to the conclusion that children access numerical representations slower than adults. But this conclusion is proved to be wrong. Temple and Posner (1998) measured event-related potentials (ERPs) during symbolic and non-symbolic number comparison task in adults and in 5-year-old children. In the symbolic task subjects had to indicate as fast as possible whether the presented number (1, 4, 6, or 9) was smaller or larger than 5. There was a considerable difference in the speed of response between children and adults: children were more than 3 times slower, lagged behind adults by approx. one second (480 ms in adults and 1495 ms in children). Surprisingly, there was no such a difference in the distance effect measured by ERPs. In fact, both children and adults showed the ERP distance effect at around 200 ms after stimulus presentation, indicating that numerical processing is indeed as fast in children as in adults. The authors suggested that the less developed response organization abilities were responsible for the delayed responses in children, while the access to numerical representations is as fast as in adults.

In a numerical Stroop paradigm (NSP), where response organization skills are supposed to be even more important than in a simple magnitude comparison task, we (Szűcs, Soltész et al., 2007) found that children's RT were significantly more influenced by the irrelevant physical attributes than adults'. Furthermore, EEG data showed that the interference effect in children was more enhanced during response initiation and response execution phase than in adults, measured by the lateralized readiness potential.

1.2. Size congruency effects in adults and in children

As it has been discussed in chapter I, the representation of numerical magnitudes is probably not encapsulated: task-irrelevant physical size, for instance, interferes with numerical decisions. Studies have shown that decisions made upon the numerical relation in the numerical version of the number Stroop paradigm (NSP) are inevitably influenced by the physical attributes of the stimuli (*size congruity effect*). For example, decisions slow down when physical size is *incongruent* with numerical magnitudes (i.e. the numerically larger number is smaller in physical size while the numerically smaller number was larger in physical size: 8 2) (Henik and Tzelgov, 1982; Tzelgov et al, 1992).

Since then, a few studies were also undertaken to explore the size congruity effect in children. Girelli et al. (2000) tested 1st, 3rd, 5th graders and adults in the NSP. In the numerical task, where subjects had to pick the numerically larger digit, they found significant size congruity effects in all age groups. More precisely, *facilitation* was significant and did not change remarkably across age groups, while there was a clear developmental pattern in the *interference* effect. *Interference* effect emerged from the 3rd grade only, suggesting that younger children were less sensitive to interference than their older peers. The authors argue that the effect of incongruity is determined by the level of integration between the two dimensions: physical size is processed automatically and inevitably by young children, but the representational integration of Arabic numerals and physical size is not fully complete yet. The link among Arabic numerals and their referents, numerical magnitudes are not yet established at 1st grade, and it develops gradually during the early years of school. Rubinsten et al. (2002) also tested 1st, 3rd and 5th graders, but they included two subgroups of 1st graders in order to reveal the possible effects of the very first year of school, when Arabic numerals are introduced to children. They found similar results regarding the size congruity effect. At the beginning of grade 1, the size congruity effect was mainly composed of *facilitation*, without the *interference* component. Interference effect became significant from the end of the 1st grade. Accordingly, in 3rd and in 5th graders we (Szűcs, Soltész et al., 2007) also found both *facilitation* and *interference* effects.

1.3. Locus of interaction: representation, decision, or response?

An interaction between two dimensions, like size and number, does not necessarily mean that these representations in fact are common or overlap. There are two different scenarios to account for the emergence of an interference effect (Henik and Tzelgov, 1982). In the first scenario, both the digit's physical size and numerical meaning are mapped onto a common magnitude representation (*representational overlap*) (e.g. Moyer and Landauer, 1967; Duncan and McFarland, 1980). In the second scenario, the representations are distinct, but they share a common decisional mechanism which runs in parallel for both dimensions and initiates two separate “subresponses” (Schwarz and Heinze, 1998). These “subresponses” can enhance the final selection of the response, when size and number are congruent with each other (I will call it *decisional overlap*). Accordingly, these “subresponses” are in conflict when size and number are incongruent, so forth this conflict needs to be resolved, lengthening the time of the final, overt decision.

With an attempt to unravel the interaction of number and size at the different processing stages, Schwarz and Heinze (1998) measured event-related potentials in a NSP. The peak latency of the P3 ERP component was used as an indicator of the termination of stimulus evaluation processes (e.g. Kutas et al., 1977). Although the P3 was not modulated by congruency, they argue that interaction happened during the stimulus evaluation phase. They came to this conclusion because the measured response-related activity did not show (either) a significant effect of congruency; there was no sign of the initiation of an incorrect response, which would indicate interference at the response level. Their argument can be criticized as being based on null findings.

Cohen Kadosh et al. (2007) also used ERPs. They found that the amplitude of the P3 ERP component was significantly modulated by congruency. They also found that the response-related ERP component was modulated by congruency, when the numerical distance between the numbers was large. They concluded that the conflict is present at the representational/decisional phases of processing, and is not resolved completely until response initiation. However, the Cohen Kadosh et al. (2007) study is not always clear, at least from the perspective of the reader. It is not straightforward from their report whether these results were found in the numerical or in the physical version of the NSP, or they the two tasks were averaged together.

We also attempted to disentangle the processing stages at which interference of numerical and physical size occurs (Szűcs, Soltész et al., 2007). We found that congruency effects were significant in the peak latency of the P300 amplitude in adults, but not in children. Furthermore, the congruency effects in adults’ response time reflected the congruency effects on their P3 ERP component, while the P3 ERP component in children did not mirror the congruency effects in their RT. We concluded that the conflict is not resolved until response initiation and execution, and in children, the conflict was more emphasized during the response phase.

1.4. Hypotheses of the present study

In the present study there were multiple purposes: one was to replicate findings on the speed of numerical processing in young children. Second was to examine the effect of a non-numerical attribute on numerical decisions, and to draw a more detailed picture on their interaction. Third, from the perspective of individual differences, I was interested whether there were any relationships among ERP measurements and behaviour. This latter is rather only an exploratory question, as I had no exact a priori hypotheses about the possible relations.

2. Methods and results

For the sake of concise representation and easier interpretation of the results, Methods and Results sections are not separated for the report of this study. Methods and results are discussed in a parallel fashion: all the different analyses are linked with their corresponding results.

Subjects. Adults were recruited via advertisements (N=25).

Children (N=20; from grades 1, 2 and 3) were recruited from primary schools in and around Cambridge, UK. All children came from middle class background and were white-Caucasian. All children had English as their first language. Written informed consent was obtained from parents and the study was approved by the Cambridge Psychology Research Ethics Committee.

2.1. Methods: Experimental paradigm and task

The numerical Stroop Paradigm (NSP) was used in this study. Participants were shown two one-digit numbers on a 17-inch screen, located approximately at 1 meter distance from their eyes. There were 4 possible number pairs, with two different numerical distances: distance 7: 1-8, 2-9, and distance 1: 1-2, 8-9. The side on which the number appeared was counterbalanced (2 instances for all number pairs). The same digits were used in both numerical distances (1, 2, 8 and 9). This was done in order to have perceptually identical digits in both numerical distance conditions, avoiding any possible perceptual effects which could have been confounded with numerical distance (and to our knowledge, this type of perceptual control has not been done in any previous studies of number comparison).

The possible instances of number-size congruency were delivered by manipulating the physical size of the presented digits. For a general overview of the varieties of number-size congruency in the NSP, please see Figure I/1.

In the *Congruent* condition, the numerically larger number was larger in physical size (50 points) than the other digit (40 points). In the *Incongruent* condition, the numerically larger digit was smaller in physical size. In the *Neutral* condition, both digits were of the same physical size (45 points) and only their numerical meaning differed. The subjects' task was to indicate the numerically larger number as fast as possible by depressing a button on a gamepad with their left or right index finger. If the number on the left was larger, they had to press the left button. Accordingly, if the number on the right was the larger one, they had to press the right button. A schematic figure of a trial is shown in Figure V/1.

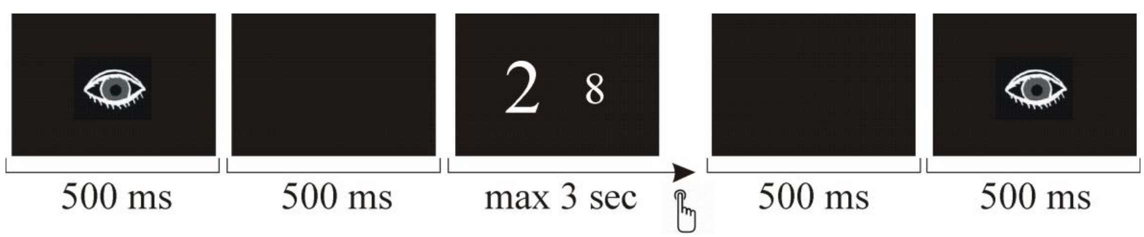


Figure V/1: Schematic presentation of a trial. The trial starts with a picture of an eye (for 500 ms), participants are asked to blink during the presentation time of these pictures if they needed. The eye is followed by a 500 ms pause, and then the stimuli are presented until a response is given, with a maximum of 3 sec. There is a 500 ms gap again before the start of the next trial.

Stimuli order was pseudo-randomized, so that all possible number pairs were presented at equal numbers and that the same stimuli could not follow more than 2 times in sequence. There were 10 blocks of 48 stimuli, preceded by 24 practice stimuli. The experiment lasted for approximately an hour[20]. Adult participants were paid £10 per hour; children were rewarded with sweets and with a custom designed T-shirt (independent of task performance).

2.2. Methods: Data collection and pre-processing

Both response time (RT, in milliseconds) and accuracy data (percent correct, %) were collected and analyzed.

Electrophysiological data were recorded by EGI's 65-channel Geodesic Sensor Net. The sampling rate was 500 Hz, an online band-pass filter of 0.01-70 Hz was used. The data was band-pass filtered between 0.01-30 Hz offline. Epochs were extracted time-locked to the stimuli, from -100 to 800 ms relative to the stimulus presentation time (where stimulus presentation time was the 0). Epochs were then baseline-corrected relative to the -100 to 0 ms interval relative to the stimulus onset. Average reference was recomputed from the Cz electrode. Epochs containing voltage deviations exceeding ±150μV relative to the baseline at any of the recording electrodes were rejected. Trials contaminated with strong alpha activity (8-10 Hz; this is especially relevant in children) were rejected: when the power of alpha exceeded 3 standard deviations from the average alpha activity within subject, trials were excluded from further analyses. Subjects showing strong alpha activity in comparison to the group average were also excluded. Noisy channels (which for example lost contact with the head surface during recording) were interpolated.

Trials only with correct responses were kept for both ERP and RT analyses. For the purpose of a one-to-one matching of ERP data and behaviour, trials excluded from ERP analyses (during artefact rejection) were not included in RT analyses either.

[20] In case of children, the number of recorded blocks was variable and depended on the actual child. We aimed for at least 5 blocks; when children seemed or indicated that they were tired, or not interested any more in the task, the experiment was halted.

2.3. Methods and results: Behavioural data

Median RT and accuracy data were subjected to a mixed design repeated measures ANOVA with the between-subject factor of Group [adults and children] and with the within-subject factors of Congruency [Neutral, Congruent and Incongruent] and Numerical distance [Distance 1 and 7]. Facilitation and interference effects are expressed via testing differences between the pairs of neutral and congruent (facilitation), and neutral and incongruent (interference) conditions. Further, measures of 'facilitation' and 'interference' were also derived and contrasted in the following way. 'Facilitation' effect was calculated by subtracting RT and accuracy in the neutral condition from RT and accuracy in the congruent condition; 'interference' was calculated by subtracting RT and accuracy in neutral condition from RT and accuracy in the incongruent condition. 'Facilitation' and 'interference' effects were compared by a separate ANOVA (Group × Facilitation/Interference) when this comparison contributed to the better understanding of congruency effects. Greenhouse-Geisser corrected p and epsilon (ε) are reported where necessary (when the sphericity assumption of the repeated measures ANOVA was violated). Post-hoc contrasts were calculated using the Tukey-Cramer test. Analyses were carried out with the Statistica (StatSoft Inc.) software.

2.3.1. Accuracy

Congruency effects in accuracy are shown in Figure V/2A.

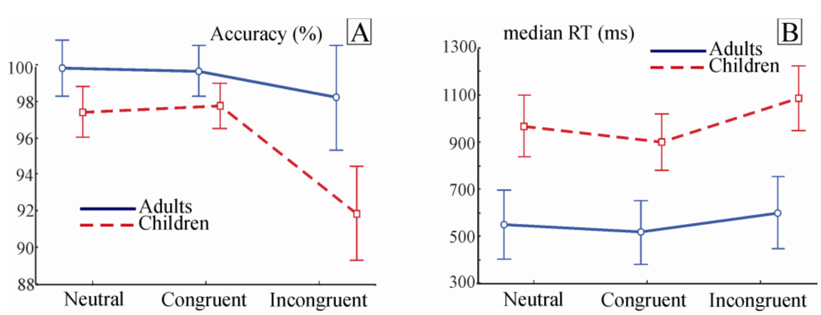


Figure V/2: Congruency effects in adults and children.

Main effects of **Group** ($F(1,43)=46.99$, $p<0.0001$) and of **Congruency** ($F(2,86)=24.66$, $\varepsilon=0.78$, $p<0.003$) were significant. Children committed more errors than adults (mean and standard error of accuracy: 95.67% (0.34) and 99.21% (0.38)). The **Group × Congruency** interaction was also significant ($F(2,86)=24.66$, $\varepsilon=0.78$, $p<0.003$). According to post-hoc comparisons, facilitation was not significant (Congruent vs. Neutral: $p=1$) in adults. Interference showed a strong statistical trend ($p=0.067$) in adults. In children, facilitation was also not significant ($p>0.9$), but interference effect was significant ($p<0.0002$) and was steeper than the interference effect in adults (see Figure V/3A).

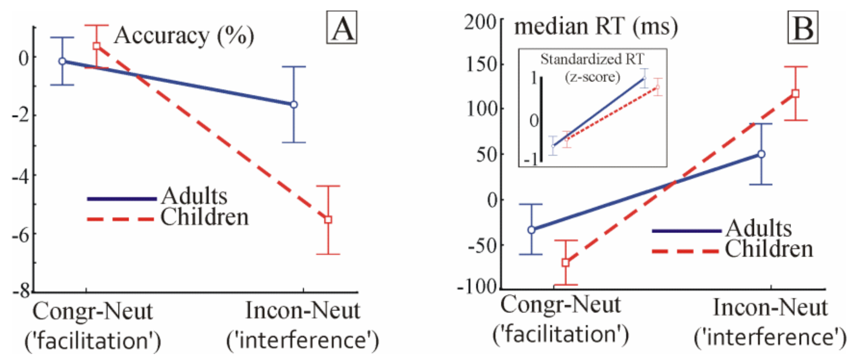


Figure V/3: 'Facilitation' (Neutral minus Congruent) and 'interference' (Neutral minus incongruent) effects in adults and children.
Insert in Figure V/3B: Facilitation and interference were calculated from normalized RTs.

The 'Interference' effect was significantly stronger in children than in adults ($p<0.0003$), while 'facilitation' effect did not differ between the two groups ($p>0.92$). The main effect of Numerical distance was not significant ($p>0.14$), but its interaction with Congruency was significant (**Congruency × Numerical distance**: $F(2,86)=4.73$, $\varepsilon=0.68$, $p<0.03$). As post-hoc comparisons revealed, numerical distance effect was significant only in the incongruent condition, in both groups (Figure V/4A). Or, in other terms, interference effect (neutral vs. incongruent) was larger in Numerical distance 1.

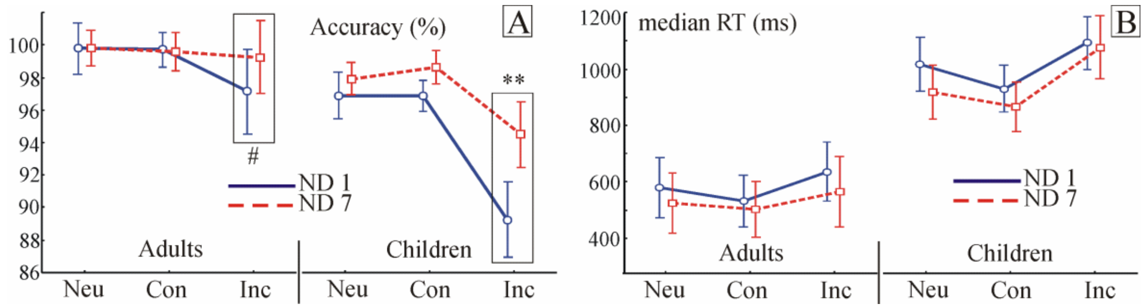


Figure V/4: Congruency × Numerical distance in both groups. # denotes p<0.62, ** denotes p<0.01.

2.3.2. Response time

Main effects of **Group** (F(1,43)=85.18, p<0.0001) and of **Congruency** (F(2,86)=76.05, ε=0.9, p<0.0001) were highly significant (Figure V/2B). Children were remarkably slower than adults (mean and standard error for children and adults: 983.64ms (30.9), 555.97ms (34.55)). Regarding Congruency effects, the **Group × Congruency** interaction was also significant (F(2,86)=11.3, ε=0.9, p<0.0001): facilitation (congruent vs. neutral) and interference (incongruent vs. neutral) effects were both significant in children (both comparisons <0.0003); in adults, only interference was significant (p<0.0002 and p>0.3). Similar to accuracy data, as can be seen in Figure V/3A, 'interference' effect was significantly larger in children than in adults (p<0.03). However, larger differences among the different conditions of congruency in children could be partly due to the generally higher values in children; aiming to correct for this, RTs were normalized (mean centered and expressed in standard deviations for each subject). 'Facilitation' and 'interference' measures were recalculated from the normalized data and entered again into a Group × Facilitation/Interference ANOVA. As insert in Figure V/3B shows, 'interference' effect is not larger any more in children than in adults (p>0.6).

The main effect of **Numerical distance** was found to be significant in RT (F(1,43)=42.01, p<0.0001). Small numerical distance yielded longer RTs (mean and standard error: 797.06 ms (54.47)) than large numerical distance (742 ms (60.81)).

2.4. Methods and results: ERP data

2.4.1. Detection of ERP effects

First, the data matrix of [electrodes × time points × participants × conditions] were entered into a point-by-point Congruency[3] × Numerical distance[2] repeated measures ANOVA, separately for both groups (children and adults). In order to deflate the possibility of the Type I error, effects were considered as significant at alpha=0.025 across at least 5 consecutive time points (10ms) and at least 3 electrode sites. Second, time points showing significant experimental effects were then averaged across the time window showing the significant effect, at each significant electrode and were entered into an Electrode × Congruency × Numerical distance ANOVA. Results of such an ANOVA are referred to as composite F and p values. Electrodes with negative and positive voltages were averaged separately and entered separate ANOVAs. The distinct analyses of electrodes with opposite polarities are necessary in order to avoid annulations of effects by averaging positive and negative values together during analysis of variance computations. Also, in some instances, although the polarities were not the opposite, the direction of the experimental effect could show the opposite pattern (the effect 'turns around'). These opposing patterns of experimental effects were revealed by post-hoc comparisons of Electrode × [experimental effect] interactions; those electrodes were also analyzed in separate blocks.

The main effect of **Congruency** was significant in three time intervals in adults (composite ANOVA results: 210-250 ms: $F(2,38)=26.24$, $\varepsilon=0.94$ $p<0.0001$; 320-420 ms: $F(2,38)=14.63$, $\varepsilon=0.74$, $p<0.0002$; and 550-700 ms: $F(2,38)=5.36$, $\varepsilon=0.96$, $p<0.01$), and in two time intervals in children (composite ANOVA results: 320-390 ms: $F(2,48)=6.67$, $\varepsilon=0.94$, $p<0.004$; and 580-680 ms: $F(2,48)=7.38$, $\varepsilon=0.87$, $p<0.003$). Topographic plots[21] indicate electrodes with significant effects and most typical waveforms are shown in Figure V/5. From the dipole-like topographic distribution of the congruency effect, voltages measured at the centro-parietal electrodes were averaged across time and were selected for the report of composite F and p values.

---

[21] spherical spline interpolations, as implemented by EEGlab software package (Delorme and Makeig, 2004) running under MATLAB

Because the **Congruency** effect found in amplitude values in the range of the **P3 ERP component** can be due to latency differences among the different congruency conditions, peak latencies of the P3 ERP component were extracted and subjected to a Congruency × Numerical distance ANOVA. A program script searched and extracted the latency of maximum amplitude in the range of 300-600 ms in adults and in the range of 400-700 ms in children. The peak latency differences among the congruency conditions were indeed significant in adults ($F(2,36)=19.89$, $\varepsilon=0.82$, $p<0.0001$) but were not significant in children ($p>019$). In adults, latency of P3 was the shortest in the congruent condition, followed by the neutral, then by the incongruent condition (mean and standard error: Congruent: 399.56 ms (51.62); Neutral: 410.48ms (57.36); and Incongruent: 429.4 ms (68.64)). Post-hoc comparisons revealed that interference was significant (neutral vs. incongruent: $p<0.002$), and facilitation showed a strong statistical trend (neutral vs. congruent: $p=0.07$).

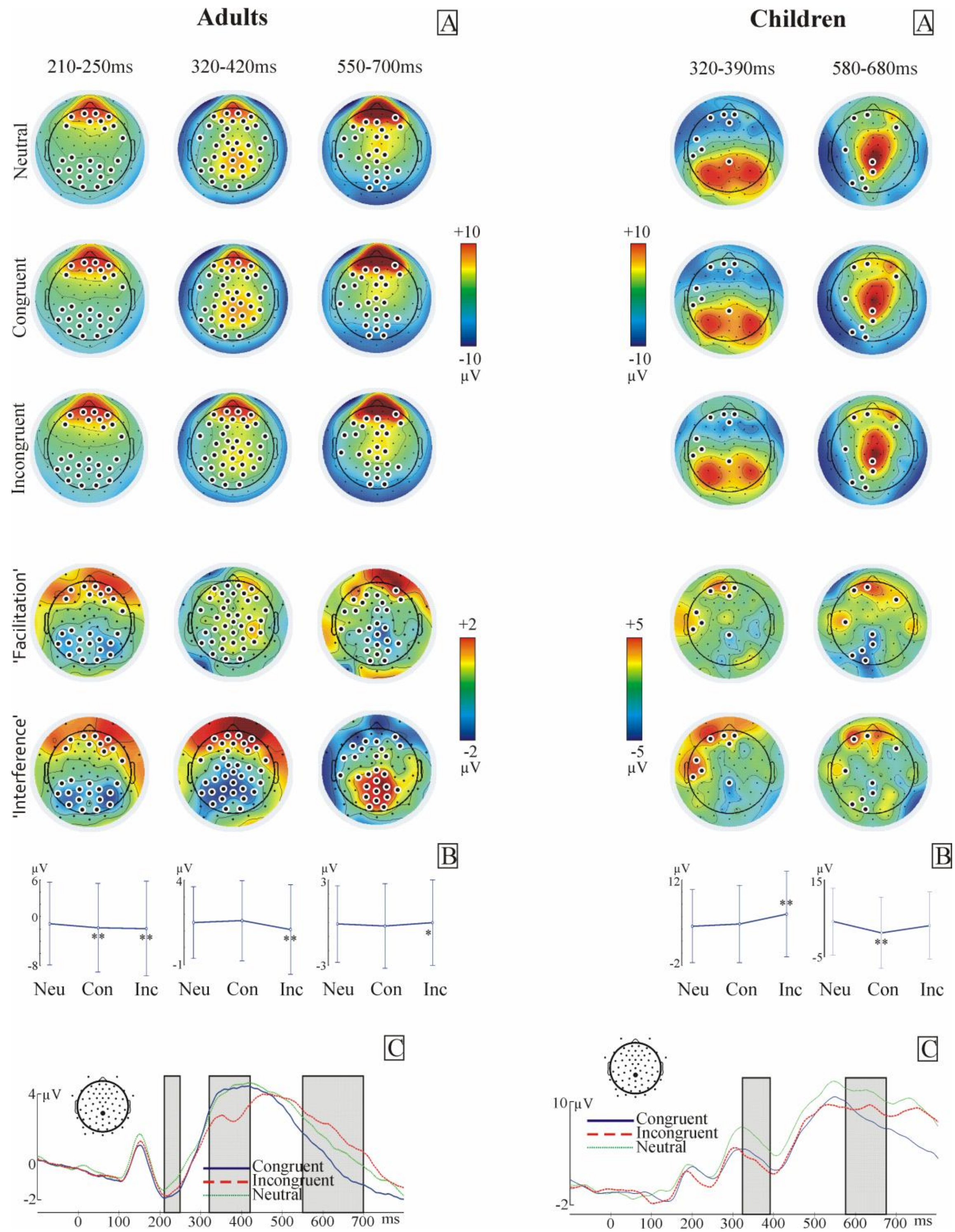


Figure V/5: A: Topographic plots of significant Congruency effects. B: Post-hoc comparisons of the Congruency conditions (Neu = neutral, Con = congruent, and Inc = incongruent), performed on amplitudes averaged over the denoted time intervals, over the centro-parietal electrodes. ** denotes p<0.01, * denotes p<0.05. C: ERPs measured on the most typical electrodes; significant time intervals are denoted by grey boxes.

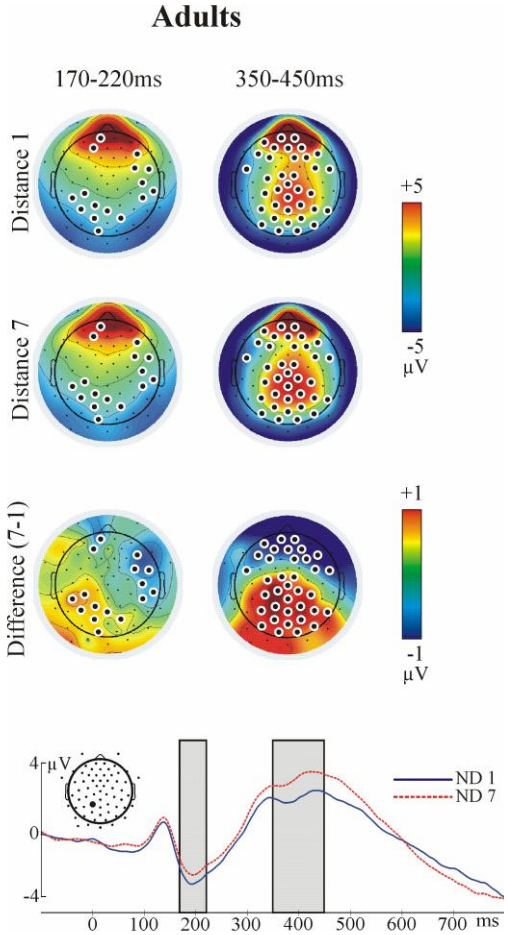


Figure V/6: Numerical distance effect in adults.

The main effect of **Numerical distance** was found to be significant only in adults (Figure V/6; Composite ANOVA results: 170-200 ms: F(1,19)=28.24, p<0.0001; 35-450 ms: F(1,19)=17.02, p<0.006) and not in children. Again, ANOVA results reported are from the ANOVA conducted on voltages over the centro-parietal sites, averaged across the time intervals showing the significant effect. The most typical waveform is also shown, with grey boxes indicating the time intervals of the significant distance effects.

The interaction of **Congruency** and **Numerical distance** was significant in both groups. In adults, two time intervals showed the significant interaction (composite ANOVA results: 320-360ms: F(2,38)=7.82, ε=0.92, p<0.002; 480-520ms: F(2,38)=8.65, ε=0.98, p<0.0009). In children, one time interval (200-300 ms) showed the significant interaction (F(2,48)=9.61, ε=0.87, p<0.0007). Figure V/7 shows topographic plots and composite results of the post-hoc analyses.

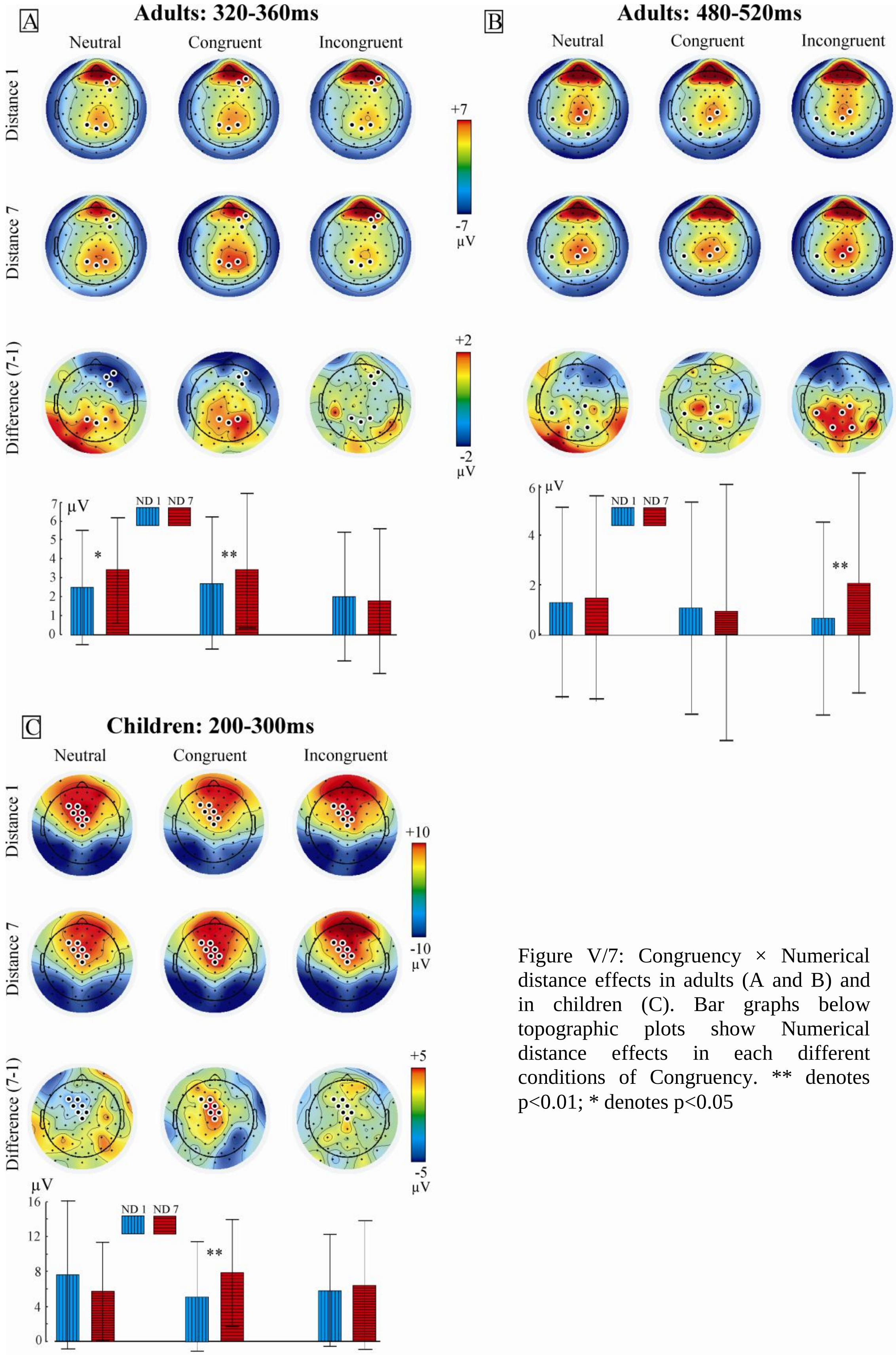


Figure V/7: Congruency × Numerical distance effects in adults (A and B) and in children (C). Bar graphs below topographic plots show Numerical distance effects in each different conditions of Congruency. ** denotes p<0.01; * denotes p<0.05

2.4.2. Correlational analyses between ERPs and behaviour

Significant effects in ERP revealed by the analyses described above were correlated with behavioural data in order to explore any functional relationship between congruency and numerical distance effects in ERPs and in behaviour. For this purpose, the root mean square (RMS) of amplitudes of the significant ERP effects were calculated and correlated with accuracy and RT effects. ERP effects of congruency were generated by the following way: neutral condition was subtracted both from the incongruent condition ('interference') and from the congruent condition ('facilitation'). ERP effects of numerical distance were calculated in a similar way: distance 1 was subtracted from distance 7. RMS was calculated with the purpose of eliminating the polarity and direction of ERP effects at different electrode sites. The absolute size of ERP effects were extracted across all significant electrodes, independently of their polarity allowing for an inclusive analysis of the congruency and numerical distance effects in ERPs.

Regarding congruency effects, no correlations emerged in either of the groups. Regarding the numerical distance effect, ERP effect size in adults did show correlations with behaviour in the following way. The smaller the numerical distance effect was in ERP between 100 and 200 ms, expressed in RMS, the faster their behavioural response was in both numerical distance conditions ($p<0.0001$ for both, Figure V/8).

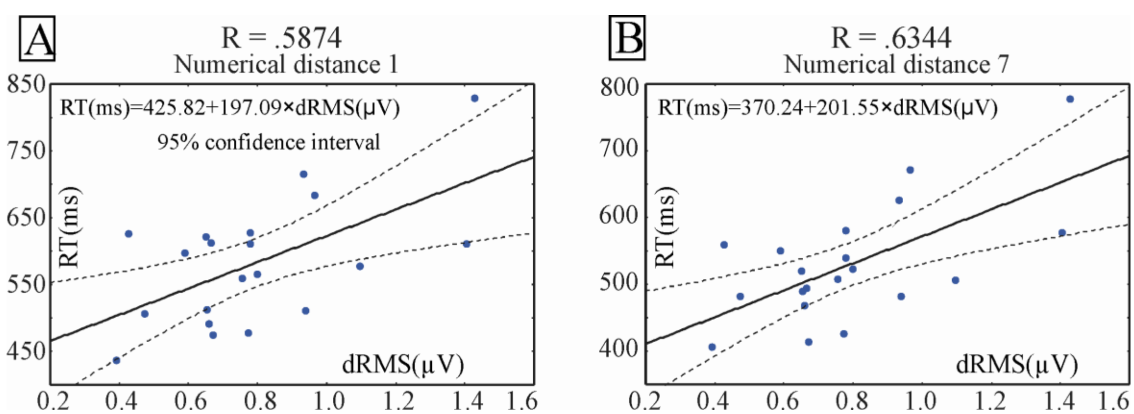


Figure V/8: Correlation of ERP distance effect (dRMS = RMS in D7 minus RMS in D1) with RT in condition D1 (A) and condition D2 (B)

As the peak latency of the P3 ERP component showed a similar pattern of congruency effects as RT did (latencies are the shortest for the congruent condition, and longest in the incongruent condition), correlations of P3 latencies with RT were also tested. Neither of the correlations was significant.

2.4.3. Comparison of ERPs between children and adults: differences in general processing speed

The peak amplitudes and the latency of peak amplitudes of major and well established ERP components were collected from all participants in both groups. Peak amplitude and latency were searched by a program script in the time windows at the specified electrodes given in Table V/1:

ERP component	Group	time window (ms)	Electrodes: Far left	left	Central	right	Far right
P1	both	50-200	28 32 36	29 33 37	30 38 39	42 41 40	46 45 44
N2	both	100-350	28 32 36	29 33 37	30 38 39	42 41 40	46 45 44
P3	adults	300-600	22 28 32	18 29 33	65 30 38	43 42 41	47 46 45
	children	400-700	22 28 32	18 29 33	65 30 38	43 42 41	47 46 45

Table V/1: time window and electrodes where peak amplitudes and latencies were searched.

Mean latencies of peak amplitudes across the selected electrodes were then compared with independent t-tests between the two groups. All three comparisons were significant: P1, N2 and P3 peaked earlier in adults than in children (sample ERP waveforms are shown in Figure V/5; mean amplitudes, t and corresponding p values are shown in Table V/2).

ERP peak	Children mean(ms)	Children SD(ms)	Adults mean(ms)	Adults SD(ms)	t	p <
P1	139.68	8.62	121.9	18.81	4.21	0.0002
N2	235.79	13.3	201.45	29.06	5.27	0.0001
P3	549.14	25.66	413.14	32.12	15.8	0.0001

Table V/2: Results of independent t-tests between the two groups, comparing ERP component latencies.

Second, peak amplitudes were entered into a Group × Location × Site × Congruency × Numerical distance ANOVA, where Location and Site refers to the topographical arrangement of the electrodes in the following way. Location comprises of: far left, left, central, right, and far right. Site refers to centro-parietal, parietal and occipito-parietal sites. The above analysis was performed in order to reveal possible hemispheric differences between adults and children. However, the ANOVA did not yield any significant interactions including Group, Location and Site ($p>0.3$ for all), suggesting that there are no differences between children and adults in the topographic distributions of P1, N2 and P3 ERP components.

2.4.4. Lateralized readiness potential

The lateralized readiness potential (LRP) indicates the initiation of motor response in the motor cortex. It is measured by comparing the electric potentials at the electrodes above the left and right motor cortex. The LRP was calculated as proposed by Coles (1989):

$$[(C4^{22}-C3)_{\text{LEFT HAND response}} + (C3-C4)_{\text{RIGHT HAND response}}] / 2$$

LRPs were subjected to a one-way ANOVA with the within-subject factor of Congruency, for both groups separately. The test did not yield any coherent or significant results.

[22] Electrode names according to the 10-20 international system of electrode placement. On the EGI's Geodesic Sensor Net, these are the no. 54 and no. 14 electrodes.

3. Discussion of the results

3.1. Numerical distance effects

Consistent with previous studies (e.g. Sekuler and Mierkiewitz, 1977; Temple and Posner, 1998; Szűcs, Soltész et al., 2007), the numerical distance effect was significant in RT. This finding confirms again that numbers are represented in a phenomenologically similar manner in children and adults. In accuracy, the distance effect interacted with congruency effect; it was significant only in the incongruent condition. Most probably the relatively high accuracy in the neutral and in the congruent conditions yielded a ceiling effect in these conditions.

In ERPs, the distance effect was highly significant in adults, already in between 170-200 ms after stimulus presentation. It is important to note here that all electrodes, and all time points were subjected to statistical analyses instead of a few predefined time windows and electrodes (like in Temple and Posner, 1998; Dehaene, 1996; Grune et al., 1993; Schwarz and Heinze, 1998; Cohen Kadosh et al., 2007). In the Szűcs, Soltész et al. (2007) we also found a significant numerical distance effect starting at around 140 ms.

Interestingly, the effect size of numerical distance in ERPs between 170 and 200 ms correlated with RT. The smaller the ERP distance effect, the faster the RT in both distance conditions. For now, I can only speculate on the exact meaning of this relationship. Distance effect decreases with age (Sekuler and Mierkiewicz, 1977; Duncan and McFarland, 1980; Holloway and Ansari, 2008), suggesting the refinement of numerical representation: the discrimination of magnitudes becomes easier and more accurate. Following this train of thought, it could be hypothesized that subjects whose numerical representation is less refined, are slower in number discrimination or number comparison tasks. But of course a third, and latent variable, like attention, can not be excluded either as an explanation for the smaller numerical distance effect and for the shorter RTs.

In children, the main effect of numerical distance was not significant; however, numerical distance interacted with the congruency effect at around 200-300 ms, and was significant at centro-parietal electrodes in the congruent condition (see Figure V/7). The timing is very similar to what we found previously (Szűcs, Soltész et al., 2007), where a significant distance effect in children was found around 210-230 ms.

3.2. Size congruency effects

In behaviour, children showed stronger *interference* effects than adults, indicating that they are more sensitive to conflicting information. This is in accord with our previous results (Szűcs, Soltész et al., 2007) but in contrast with the findings of Girelli et al. (2000) and Rubinsten et al. (2002), where facilitation was significant and interference emerged with age. This contradiction among these studies has to be resolved in the future. Meanwhile, regarding the present results, it seems reasonable to argue that the irrelevant physical size of the stimuli interferes with the relevant numerical dimension, indicating that the processing of numerical magnitudes is *not encapsulated* from other physical magnitudes.

In ERPs, the effect of congruency was significant at three different time window in adults and in two different time window in children. Adults recognized when stimuli were not neutral and varied in physical size (see Figure V65B, first column), and probably invested more attentional effort in these cases in order to eliminate the interfering information of physical size. In children, the first component (or time-interval) did not show significant interference effect, probably reflecting a less efficient attentional control in children than in adults. *Interference* being stronger in RT and accuracy could also reflect the weaker attentional and inhibitory control abilities in children and it is also in line with other studies showing that children are more susceptible to interference than adults (Bunge et al., 2002; Durston et al., 2002; Schroeter et al., 2004).

Meanwhile, according to the framework proposed by Posner (1978), *facilitation* effect indicates the automatic nature of processing; physical size is inevitably evaluated, even when it is irrelevant to the task. The lack of facilitation in the behavioural data of adults can be explained by ceiling effects; they were 100% accurate and almost equally fast in the neutral and congruent condition. However, the latency of the P3 ERP component showed a strong effect of facilitation, suggesting that stimulus evaluation and/or decision was in fact enhanced in adults in the congruent condition.

3.3. The locus of interaction

In adults, the numerical distance effect arose earlier than and independent of the interfering effects of the irrelevant physical dimension. In children, this is not the case: the effect of distance was modulated by congruency effects. These findings suggest that adults process numerical magnitudes somewhat earlier than children (approx 50 ms).

Regarding the question whether the interactions between number and physical size happen at the level of magnitude representation (*representational overlap*), or rather it were only a phenomenon caused by the common mental measure used for both magnitudes (*decisional overlap*), I think can not be answered by this paradigm. However, the early appearance of the distance effect is suggestive and might indicate the access to the representation of numerical magnitudes. Although this assumption is not completely supported by *priming experiments* where the prime stimuli elicited numerical distance effects (targets preceded by a close prime were responded faster: Heyer and Bryant, 1986; Dehaene et al., 1998; Naccache and Dehaene, 2001b; Reynvoet and Brysbaert, 2004; Reynvoet et al., 2002). Recently, Reynvoet et al. (2009) investigated the priming distance effect in children. In a number comparison task children were asked to indicate whether the presented number was smaller or larger than 5. It was found that the priming distance effect was significant already in 1^{st} grade children: targets preceded by a close prime (e.g. 3 preceded 4) were responded faster than targets preceded by a farer prime (e.g. 1 and 4). Meanwhile, they also found that the priming distance effect did not correlate with the distance effect in the comparison task. They concluded that priming distance effect operates at the level of representation, and the comparison distance effect operates at the level of comparison mechanisms.

If we accept that the numerical distance effect emerges at the decisional level, from the results of the numerical Stroop experiment I can conclude that in adults, the comparison of numerals happens somewhat earlier than the unavoidable and unintentional comparison of physical sizes.

In children, the picture is more complicated. There was no numerical distance effect independent of congruency effects. This result could be explained in two different ways. First, the representation/decision of numerical magnitudes overlaps with the representation/decision of physical magnitudes in children, suggesting that perceptual information still predominates symbolic information in children (Piaget, 1952). Second, the overlap is due to the immature attentional and inhibition processes in children. In the future, paradigms which are able to disentangle representation, attention and decision may reveal the dynamics of these processes.

4. Conclusion

In summary, it was found that children are almost as fast in processing numerical information as adults, measured by the numerical distance effect in ERPs. However, this delay does not seem to be domain-specific, as all (not task-specific) ERP components peaked somewhat later in children than in adults, suggesting a little slower general processing speed in children. Furthermore, conflicting information from task-irrelevant properties of the stimuli exerted greater influence on children's performance, confirming previous findings on the weaker attentional and inhibitional abilities in children.

VI. A spatiotemporal principal component analysis of processing stages in the number Stroop paradigm

Abstract

The numerical Stroop paradigm has been successfully used to measure the speed and automaticity of number processing in adults and in children, using ERP methods (see chapter V**;** see also Szűcs, Soltész et al., 2007). The following study is a methodologically extended, explorative post-hoc analysis of the previously presented ERP data. Spatiotemporal principal component analysis (PCA) was used in order to reduce data dimensionality and to reveal underlying *functionally* relevant components which are otherwise concealed in ERPs. PCA was implemented using a custom-written toolbox, as it is described in chapter II/2. Statistical analyses of the PCA results were also done by using custom-written program codes. The P3 ERP component has been successfully broken down into two separate principal components, showing qualitatively distinct experimental effects. Further studies are clearly required to establish the validity of the present method; meanwhile, we claim that the PCA followed by a point-by-point ANOVA approach provides meaningful and valuable complementary information to ERPs.

1. Rationale and background

1.1. The P3 ERP component: an introduction

A long positive deflection in ERP starting at around 300 ms after stimulus presentation, first described by Sutton et al., (1965), is referred to as P3 ERP component. Sutton et al. (1965) interpreted the P3 as a component reflecting the delivery of information to the subject. Later, it has been described as an electrophysiological reflection of a mechanism that updates the content of working memory according to changes in the environment (Donchin, 1981, Donchin and Coles, 1988). The amplitude of the P3 is a function of subjective expectancy, and it is also the function of the relevance of the given stimulus in the light of task requirements. The more unexpected the stimulus is, the larger the amplitude. Also, the more relevant the stimulus is for the task, the larger the amplitude of the P3. Its peak latency is dependent on the complexity of the stimulus and on the time required for stimulus evaluation and stimulus categorization (Kutas et al., 1977). Although the P3 was traditionally manipulated and tested in oddball paradigms (e.g. Donchin, 1981, Spencer et al., 2001), it has also been used as a general measure of cognitive processing in several experiments across a broad range of cognitive research (for example, in visual emotion processing: Pourtois et al., 2008; and in numerical processing: Grune et al., 1993, Dehaene, 1996, Szűcs, Soltész et al., 2007). Because the peak latency of the P3 varies in function of the time required for stimulus evaluation, meanwhile it is relatively independent of response selection and execution (McCarthy and Donchin, 1981; Magliero et al., 1984), the latency of P3 is often used as a measure of stimulus evaluation in time.

1.2. The P3 ERP component: number processing

In the field of numerical processing, it has been found that the latency and amplitude of P3 is sensitive to the symbolic numerical distance (Grune et al., 1993; Dehaene, 1996), during number comparison tasks. With increasing numerical distance, the amplitude of the P3 increases. The amplitude of P3 was most probably related to decision making in this case: the response was selected with more confidence in the large numerical distance condition. In the small numerical distance condition, the decision was made less confidently and probably with larger variability in time, smearing the peak of the P3 on the average waveform.

In number Stroop tasks, the latency of the P3 was found to be sensitive to both facilitation and interference, both in adults and in 9-11 year-old children (Szűcs, Soltész et al., 2007). In sum, the changes of the P3 ERP component in function of the numerical distance reflect the difficulty of the task, just as the numerical distance effect observed in response times. The larger the numerical distance, the easier, so forth the faster the evaluation of stimuli and the decision made upon the stimuli. Similarly, in the number

Stroop task, task-relevant decisions are made easier and faster under congruent than under incongruent condition.

1.3. P3 ERP component: functionally different components overlapping in time

The P3 is a relatively broad component in time. So forth it is reasonable to assume that it is a composite of more than one underlying cognitive components. Accordingly, the P3 can be broken down into at least two *functionally* separate underlying components: into the 'Novelty P3' or with its other label the 'P3a', and into the 'P3b' (Ritter et al., 1968; Spencer et al., 1999b). Spencer et al. (1999a; 2001) used PCA to distinguish among the overlapping constituents of the P3. In a modified oddball paradigm, Spencer et al. (1999a) found that attention, task relevance and salience of the stimuli influence the manifestation of Novelty P3 and P3b in different ways. When the oddball sequence was attended, target (or task relevant) deviants elicited large P3 and small Novelty P3. Meanwhile, highly salient, but non-target deviants elicited both large P3 and large Novelty P3. The authors concluded that the P3 and the Novelty P3 are distinct components, and it is the summation of this two underlying components what is measured by conventional ERP methods. They also note that the so-called P3a measured in unattended conditions (e.g. Squires et al., 1975) is also a summation in ERPs of a small Novelty P3 and of a small P3, so forth it is not simply a 'shifted' version of P3.

Taking these findings together, we can tell that the P3 ERP component is an important measure of cognitive processing. It is related to the speed of stimulus evaluation, and it has been found to be sensitive to symbolic numerical information, so forth providing a reliable tool for the examination of automatic symbolic number processing.

In the present study we set out to investigate the possible underlying components of the P3 ERP component during number processing. Our working hypothesis was that if P3 is a composite of at least two, *functionally* different underlying components, probably these two underlying components are distinguishable during NSP as well. Besides our main question, the functionality of PCA methods was also addressed: how, and with what can PCA contribute to traditional ERP analyses?

2. Methods and results

2.1. Methods: Experimental paradigm and task

The subjects are identical with the adult participants of the study presented in chapter V. Children's ERPs were also analyzed, however, due to the high level of noise the PCA did not yield any coherent results (for more details please see Appendix at the end of this chapter).

The numerical version of the Numerical Stroop Paradigm (NSP) was used in this study, as presented in Chapter V.

2.2. Methods: Data collection

Electrophysiological data collection is described in detail in chapter V.

2.3. Methods and results: P3 ERP component

The peak amplitude and peak latency of the P3 ERP component were collected in both children and adults by an automatized program script, as described in chapter V (2.4.3.).

For both groups, the peak amplitude and latency of the P3 ERP component were entered into a Congruency × Numerical distance ANOVA.

2.3.1. P3 ERP component - Adults

In adults, the *amplitude* of the P3 ERP component showed both significant **Congruency** ($F(2,38)=8.17$, $\varepsilon=0.98$, $p<0.002$) and **Numerical distance** effects ($F(1,19)=47.54$, $p<0.0001$). Regarding the congruency effect, pots-hoc comparisons revealed significant interference effect (neutral vs. incongruent: $p<0.03$), but no facilitation effect ($p>0.4$). Regarding the numerical distance effect, the amplitude was more positive in the large distance condition (Distance 1: 3.58μV, SE: 5.17. Distance 7: 4.33 μV, SE: 5.44).

The *latency* of the P3 ERP component was significantly modulated by **Congruency** ($F(2,38)=19.81$, $\varepsilon=0.81$, $p<0.0001$). Again, interference was significant ($p<0.002$), and facilitation showed a strong statistical trend ($p=0.07$).

2.3.2. P3 ERP component - Children

In children, neither Congruency nor Numerical distance was significant in the *amplitude* of the P3 ERP component (p>0.5 for both).

In the *latency* data, the **Numerical distance** was almost significant (p=0.051). The P3 peaked earlier in the large numerical distance than in the small numerical distance (Distance 1: 532.98µV, SE: 45.18. Distance 7: 525.53, SE: 41.57).

2.4. Methods and results: Spatiotemporal principal components analysis (PCA)

2.4.1. Methods: Spatiotemporal PCA and statistical analysis

In the present study, the relatively high spatial resolution of the acquired data (64-channel dense array Geodesic Sensor Net) allowed for a spatial, then for a temporal PCA. PCA was performed on the covariance matrix by a custom developed toolbox using MatLab 7.1 SVD algorithm for singular value decomposition. Detailed description of PCA calculations are presented in chapter II/2.

First, a spatial PCA was performed: ERP data matrix of time points (450 [-100 to +800ms relative to stimulus presentation, sampled at 500Hz]), electrodes (65), subjects (20) and conditions (6) was first transformed into a matrix where columns represent variables (electrodes) and rows represent observations (time points, subjects and conditions). Six spatial components were selected using the Scree test (Cattell, 1966), explaining 92.2% of the variance in the original data. Original ERPs and extracted spatial components (SC) are shown in Figure VI/1 and in Fig. VI/2.

In Figure VI/2, topographic plots represent SC loadings or ’virtual electrodes’; time series of µV values represent SC scores, the time courses of SC loadings, or ’virtual ERPs’ (Spencer et al., 2001). SC loadings are correlations (from -1 to +1) between the given SC and the raw data at each electrode. SC scores are on a µV scale and represent the difference of the predicted data from the mean of the original data. SC scores are calculated by projecting the original data (with 65 channels) into the new vector space of 6 virtual channels. Signs (±) of SC loadings are arbitrary, out of the nature of the singular value decomposition (SVD).

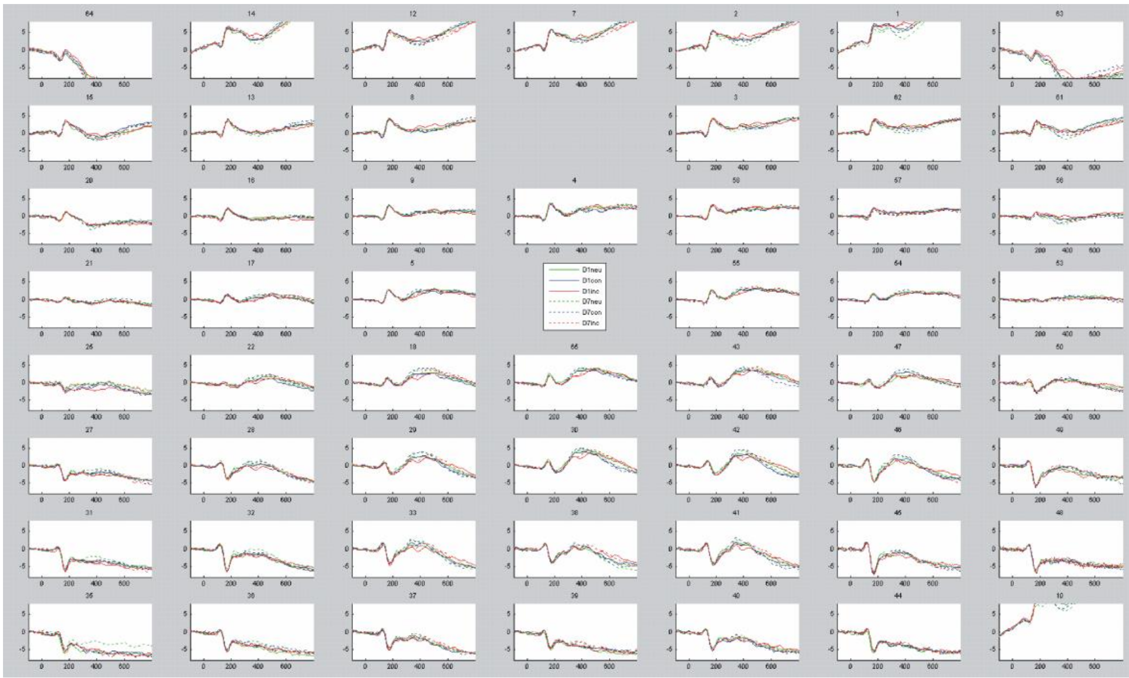


Figure VI/1: Average ERPs (N=20). Electrode plots are in topographic arrangement (approximately). The ERP data were first decomposed into spatial principal components (see below).

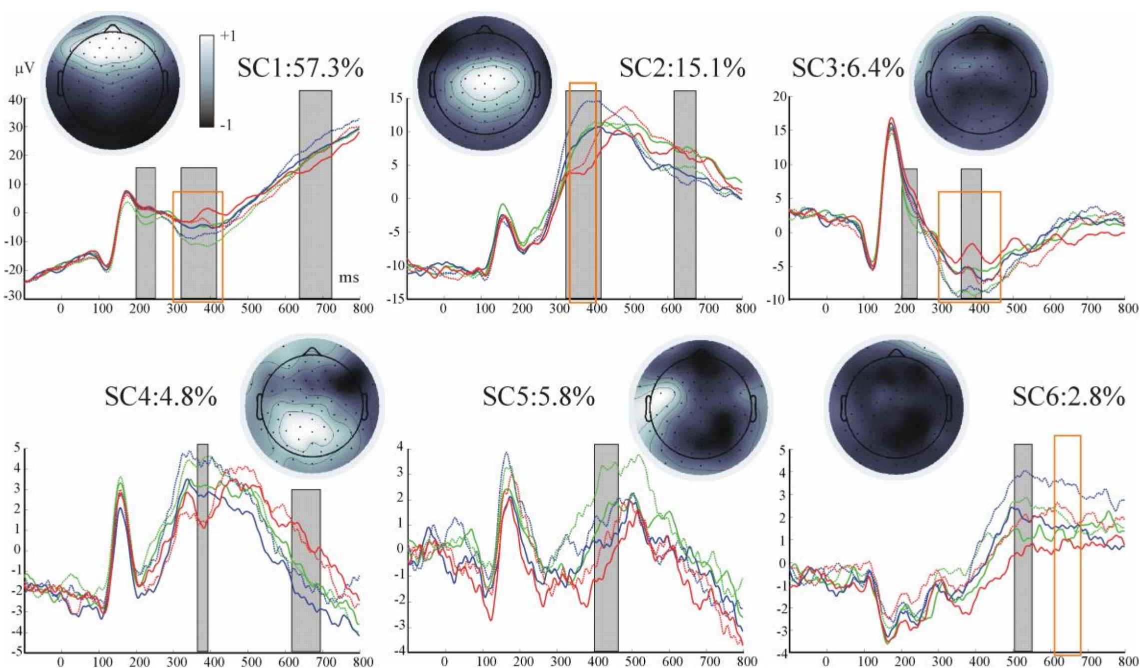


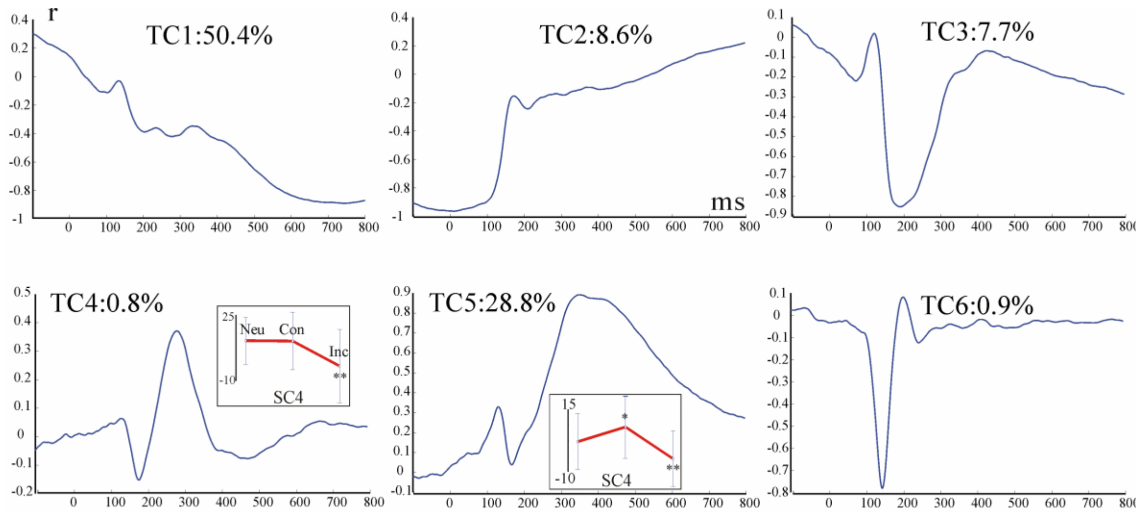
Figure VI/2: Varimax rotated spatial component (SC) loadings and the virtual ERPs belonging to each SC. SC topographic plots represent correlations of SCs with the original data, in space. The sign of correlation is random. Recomposed component scores plotted as time series for each condition, are on a μV scale (virtual ERPs), and are also random in direction. Each component explains certain variance of the original data, noted on each figure.

Significant Congruency effects are denoted by grey boxes. Numerical distance effects (or interactions with Congruency) are denoted by orange outlines. Colour codes in virtual ERPs: green: neutral, blue: congruent, red: incongruent. Solid line: numerical distance 1. Dotted line: numerical distance 5.

SC1 reflects activity on the frontal electrodes (as can be seen if activity on the frontal electrodes and activity on SC1 are compared, in Figure VI/1 and VI/2). SC2 reflects activity on central electrodes, and its time course is reminiscent to the centro-parietal P3 ERP component. SC3 represents activity measured on the bilateral parietal electrodes, together with activity from frontal locations. The identification of SC3 is ambiguous: it might represent a principal component otherwise concealed in ERP, so forth not known for the ERP literature; or can be a 'leakage' of variance from one principal component, without an independent portion of explained variance (Chapman and McCrary, 1995, see chapter II/2). One way to identify principal a component with a functional component of cognitive processing is to test its variance in function of experimental manipulations. The same argument stands for SC4, SC5 and SC6, as they are all similar to SC3, with less explained variance of the original ERP data.

Second, virtual ERPs calculated from spatial components were entered into a point-by-point ANOVA, with the within-subject factors Congruency and Numerical distance. The point-by-point ANOVA is conducted in order to reveal time intervals on each SCs showing significant experimental effects. The scale of spatial components and time intervals of interest will be narrowed, without a priori definitions of regions of interests. In order to deflate the possibility of the type I hazard, the alpha level was set to 0.01, and at least 5 consecutive time points (10ms) had to show significant experimental effects. Time intervals on SCs showing significant experimental effects are shown in Figure VI/2.

Third, the component scores of the spatial PCA ('virtual electrodes') were subjected to a temporal PCA with the variables of time points (450) and observations of 'virtual electrodes' (6), subjects (20) and conditions (6). 97.9% of the variance in the raw data is explained by the 7 temporal components (TC, or 'virtual epochs') selected for data re-composition (Figure VI/3). TC1 is a very typical component found in PCA studies, and it is most probably related to the baseline (Spencer et al., 2001). TC2 speculatively could be related to processes like expectation and preparation before and around the stimuli (Chapman and McCrary, 1995). TC3, TC4 and TC5 are easier to identify with 'classical' ERP components: N2, P3a and P3b. Again, instead of their pattern over time, their *variance in function of experimental manipulations* will allow us to make firm conclusions about these components.



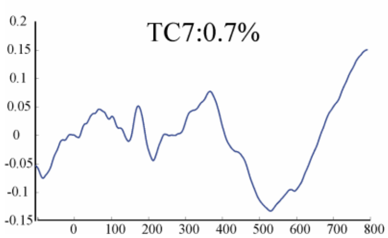


Figure VI/3: Varimax rotated temporal component (TC) loadings of the SC scores. SC scores were obtained from a sPCA, and were entered into a tPCA. TC waveforms represent correlations (R) of the given TC with the original data. Please note the different scales of Y-axes. Explained variance by each TCs is noted.
Inserts in TC4 and TC5 show the pattern of Congruency effects according to post-hoc comparisons. ** denote p<0.01; * denote p<0.05.

Fourth, factor scores were calculated for each SC combined with all possible TCs, yielding one score per one SC-TC combination. Virtually, these scores reflect the actual ‘strength’ of a TC at a given SC.

Fifth, a point-by-point ANOVA was run on the data matrix of virtual electrodes (SC), virtual epochs (TC), subjects and conditions, with the within-subject factors of Congruency and Numerical distance. Again, in order to avoid false positives arising from multiple comparisons, alpha was set to 0.01. Tukey-Kramer post-hoc comparisons of the levels of Congruency were done with Statistica software.

2.4.2. Methods: Correlations of principal components and behaviour

In a further attempt to reveal functional relations among principal components and overt behavioural responses, TC-SC combinations with significant experimental effects were subjected to correlational analyses with the behavioural data. Factor scores were grouped according to experimental conditions (e.g. congruent, incongruent and neutral) for each TC-SC combinations and were correlated with the corresponding accuracy and response time data.

2.4.3. Results: Spatiotemporal PCA, Congruency effects

The point-by-point ANOVA resulted in a matrix of significances, for all possible CS-TC combinations (Table VI/1) for the Congruency effect.

		TC1	TC2	TC3	TC4	TC5	TC6	TC7
SC1	***F***	2.2723	**8.8359**	**9.1111**	**22.097**	1.4363	1.2015	**7.1079**
	p	0.1169	**0.0007**	**0.0006**	**0.0000**	0.2504	0.3119	**0.0024**
SC2	***F***	1.0575	**5.8872**	5.1597	**13.759**	**7.9844**	2.2729	**12.886**
	p	0.3573	**0.0059**	0.0104	**0.0000**	**0.0013**	0.1169	**0.0001**
SC3	***F***	0.2125	**5.9787**	4.4327	**15.992**	0.7400	4.5448	3.4147
	p	0.8095	**0.0055**	0.0186	**0.0000**	0.4839	0.0170	0.0433
SC4	***F***	1.0573	**6.4288**	2.7871	**11.735**	**8.0713**	1.9899	**9.5553**
	p	0.3574	**0.0039**	0.0742	**0.0001**	**0.0012**	0.1507	**0.0004**
SC5	***F***	3.2103	3.3876	0.2481	0.6017	0.8797	1.3569	1.1247
	p	0.0515	0.0443	0.7815	0.5530	0.4232	0.2696	0.3353
SC6	***F***	1.2698	0.5421	4.0765	1.2996	**8.5359**	1.6252	3.2989
	p	0.2925	0.5860	0.0249	0.2845	**0.0009**	0.2103	0.0477

Table VI/1: Congruency effects (F and p values) on SC-TC combinations. Alpha was set to 0.01. Significant effects are denoted by black bold typesetting.

Temporal components (TC) showing significant experimental effects and the associated spatial components will be discussed, with an attempt to identify these principal components with meaningful stages of cognitive processing. Ideally, a TC showing significant experimental effects would be related to one corresponding SC; clearly, this is not the case. Instead, more TCs are associated with more SCs according to the ANOVA (Table VI/1). By definition of PCA, one might expect that the components are all independent from each other and all show categorically distinctive pattern. However, previous studies using PCA (e.g. Spencer et al., 2001) also associated more than one temporal factor with more than one spatial factor, even without a systematic analysis of variance of each factor. The careful examination of experimental effects and the size of these effects will help us to identify the relevant and meaningful TC-SC combinations.

TC2 is difficult to interpret, although it showed significant effects of Congruency manipulations. We would like to set aside this temporal component for now as the current experimental paradigm does not allow for firm assumptions about this component.

TC3, the temporal component with high loadings at around 200ms, shows significant Congruency effects at 'virtual electrode' 1. The Congruency effect on this TC-SC combination can be identified with the ERP congruency effect between 210-250ms (see Figure VI/1), although only interference reached significance ($p<0.0005$). Topographic distribution of ERPs at this time interval consisted of a positive peak over frontal and a negative peak over parietal electrodes. Negativity around 200ms after stimulus onset at parietal electrodes is often linked to spatial attention (Luck and Hillyard, 1994, Luck et al., 1997, Eimer, 1996) and to some aspects of working memory maintenance (Vogel and Machizawa, 2004).

TC4, with high loadings at 300ms, was found to be associated with SC1, SC2, SC3, and SC4. Post-hoc Tukey-Cramer tests revealed significant interference effects on TC4 at all four 'virtual electrodes' ($p<0.005$ for all comparisons, see insert in Figure VI/3, at TC4). Remarkably, TC4 showed interference, but no facilitation effect ($p>0.3$) at all four 'virtual electrodes'.

TC5, with its time course reminiscent to P3b, is associated with central and parietal topographic distributions (SC2, SC5 and SC6, see Table VI/1). Remarkably, TC5 shows a quantitatively different pattern of Congruency effects than TC4 (see insert in Figure VI/3, at TC5). According to post-hoc analysis, TC5 shows both facilitation ($p<0.03$) and interference ($p<0.008$) effects.

TC7, with its high loadings towards the end of the time epoch (~400-700ms), also showed significant Congruency effects. TC7 was found to be associated with three 'virtual electrodes' (SC1, SC2 and SC4). The pattern of Congruency effect is similar to the Congruency effect on TC5, showing significant facilitation (p<0.04), and weaker interference (significant at SC2, but not at the other two spatial components).

2.4.4. Results: Spatiotemporal PCA, Numerical distance effects

Table VI/2 shows Numerical distance effects (F and corresponding p values) across all SC-TC combinations.

		TC1	TC2	TC3	TC4	TC5	TC6	TC7
SC1	***F***	0.4369	**30.608**	**18.799**	**27.934**	4.7527	4.5659	0.1469
	p	0.5166	**0.0000**	**0.0004**	**0.0000**	0.0420	0.0458	0.7057
SC2	***F***	0.0002	7.7281	**12.160**	**13.973**	0.0016	7.1246	1.4779
	p	0.9882	0.0119	**0.0025**	**0.0014**	0.9688	0.0152	0.2390
SC3	***F***	0.8258	**37.351**	**19.876**	**17.390**	3.442	2.1496	0.0459
	p	ő0.3749	**0.0000**	**0.0003**	**0.0005**	0.0791	0.1590	0.8326
SC4	***F***	0.4449	3.3197	0.3362	1.9430	0.2429	0.0581	0.2612
	p	0.5128	0.0842	0.5688	0.1794	0.6278	0.8122	0.6152
SC5	***F***	0.4705	6.8057	0.0177	4.3564	4.6947	0.1038	0.5440
	p	0.5011	0.0173	0.8955	0.0506	0.0432	0.7508	0.4698
SC6	***F***	7.0459	4.2973	0.2676	0.4033	6.2529	0.0370	3.2870
	p	0.0157	0.0520	0.6109	0.5330	0.0217	0.8496	0.0857

Table VI/2: Numerical distance effects (F and p values) on SC-TC combinations. Alpha was set to 0.01. Significant effects are denoted by black bold typesetting.

Numerical distance effects at **TC3** and **TC4,** associated with SC1 and SC3 (consecutively), most probably reflect ERP numerical distance effects at around 200ms and around 300-400ms (see chapter V - Figure V/6) components. Unlike the effect of congruency, numerical distance effect can not take qualitatively different forms, which could be disentangled by PCA. Unfortunately, the present paradigm does not include appropriate manipulations for the differentiation of the various SC-TC combinations showing the numerical distance effect.

3. Discussion of the results

3.1. Principal components underlying the P3 ERP component

Results of the present PCA revealed *functionally* different, independent principal components in the time range of the P3 ERP component during number processing. First, a principal component being active at around 300 ms, was found to be sensitive to incongruence, but not to facilitation. This pattern probably reflects conflict detection or resolution. Second, a principal component with a broader time course in the range of approx. 300-600 ms was sensitive to both facilitation and interference, showing a qualitatively different pattern of congruency. This latter principal component could presumably be associated with decision processes. But how can we reconciliate these principal components with the ERP results? We have found that congruency expressed a significant effect on the P3 ERP component (in adults). Notably, in the *latency* of the P3 ERP component, interference was significant but facilitation also showed a strong statistical trend; this pattern could be explained by the two underlying principal components in the following way. One principal component, loading the most around 300 ms, showed a strong interference, while the other principal component, loading at a later time interval, showed both facilitation and interference. In the P3 ERP component, the temporal and spatial summation of these two principal components is reflected; there is a strong interference, and a somewhat weaker facilitation.

In the future, besides the replication of the present results, planned manipulation of the two separate principal components are required.

3.2. PCA: Methodological considerations

In regard to PCA methods, the results are not straightforward. Here we cannot make firm conclusions about the several principal components with significant experimental effects. Further examination of the revealed components is necessary, with the utilization of more targeted experimental designs. The multiple associations of temporal components and spatial components resulted from variance analyses raise the question whether the thoughtful judgement of the human researcher would be still unavoidable for the selection of relevant components. However, an attempt has been made for the establishment of an analysis route which relies merely on statistical computations, and, ideally, does not require any judgements of the researcher.

4. Conclusion

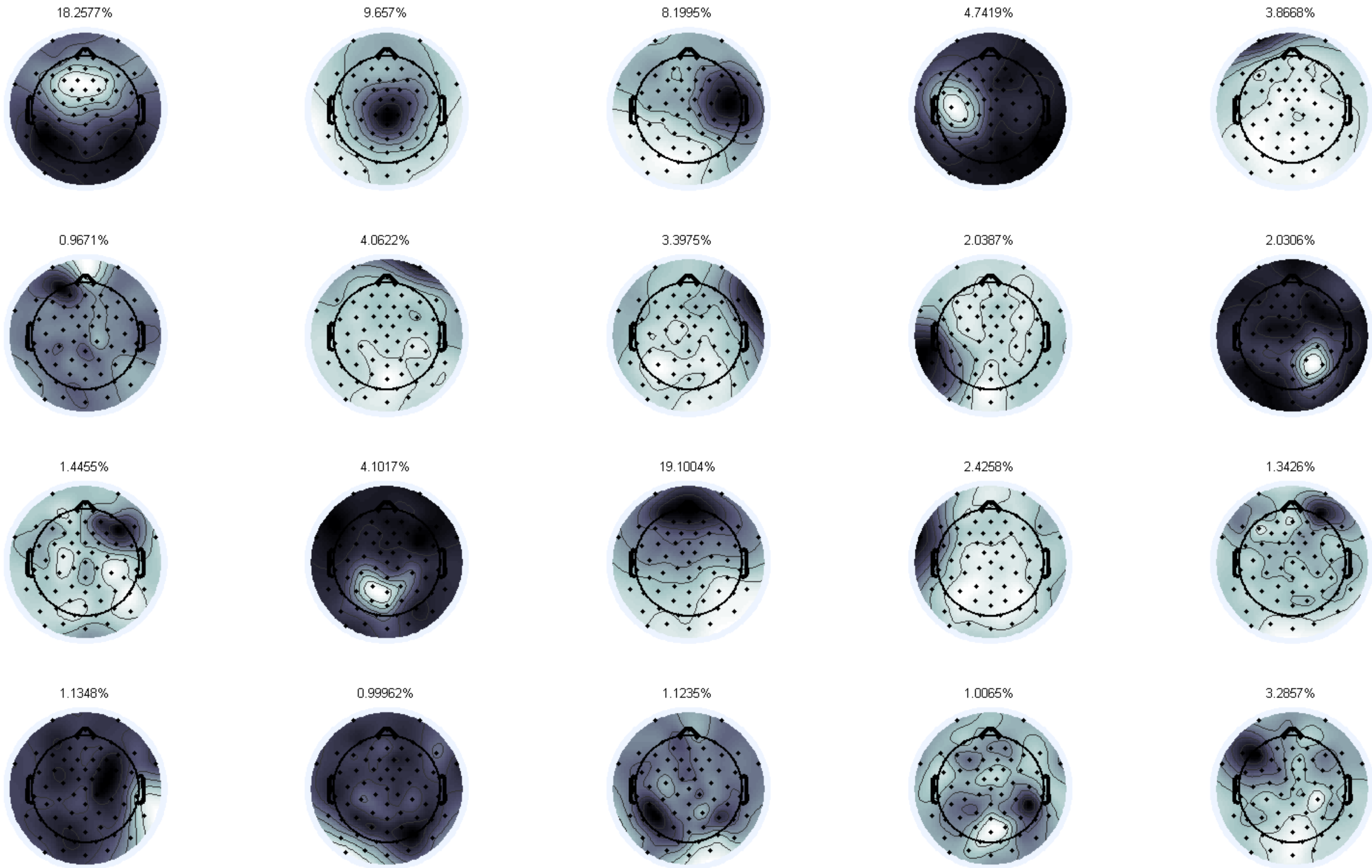
In the present study a spatiotemporal PCA has been conducted on ERPs recorded during the numerical Stroop paradigm in adults. The analysis was motivated by previous findings, showing that the P3 ERP component can be broken down into *functionally* different components of cognitive processing. These underlying components overlap in space and time, so forth are not distinguishable by the traditional ERP measures. The P3 ERP component is relevant for us for multiple reasons. First, it is a well established and replicable marker of the time required for stimulus evaluation and categorization. Second, it has also been used in studies investigating the speed and automaticity of numerical processing in adults (Grune et al., 1993; Dehaene, 1996; Szűcs, Soltész et al, 2007), in children (Szűcs, Soltész et al., 2007) and also in adolescents with developmental dyscalculia (Soltész et al., 2007).

We found that the two principal components underlying the P3 ERP component in fact showed qualitatively different patterns of congruency effects: the earlier principal component, with the highest loadings at around 300 ms, was sensitive to interference, but not to facilitation. Meanwhile, a somewhat later principal component was sensitive to both interference and facilitation. These results are important in how one interprets the *functionality* of the P3 in a numerical Stroop paradigm.

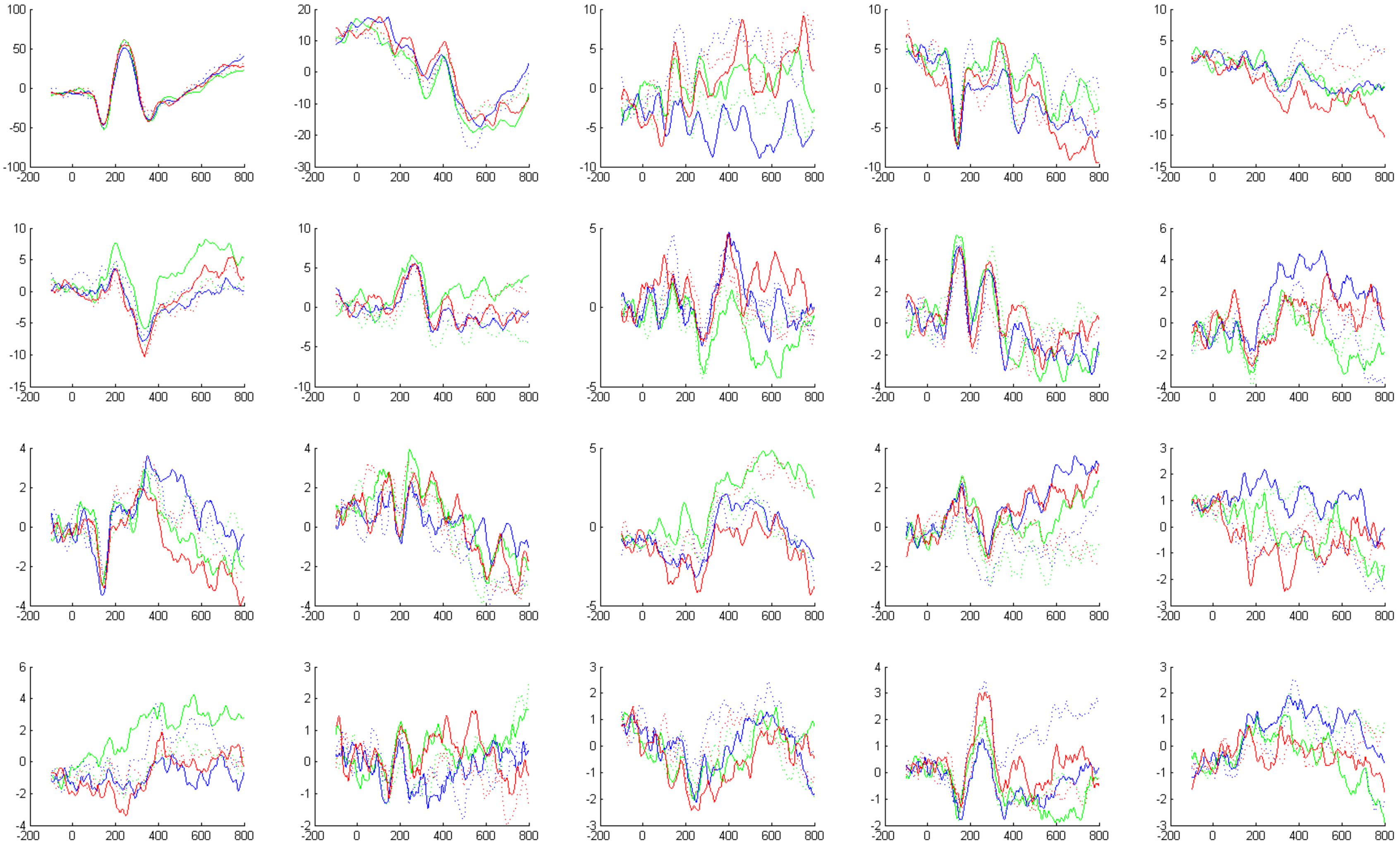
5. Appendix: Children's data

Children's ERPs were also subjected to spatiotemporal PCA; however, due to the large (clearly visible) noise and between-subject variability, peculiarities of recordings from children, the results of the PCA contained several ambiguous principal components and were highly inconsistent (Appendix VIa/1 A, B, and C shows the spatial and temporal factors for illustration).

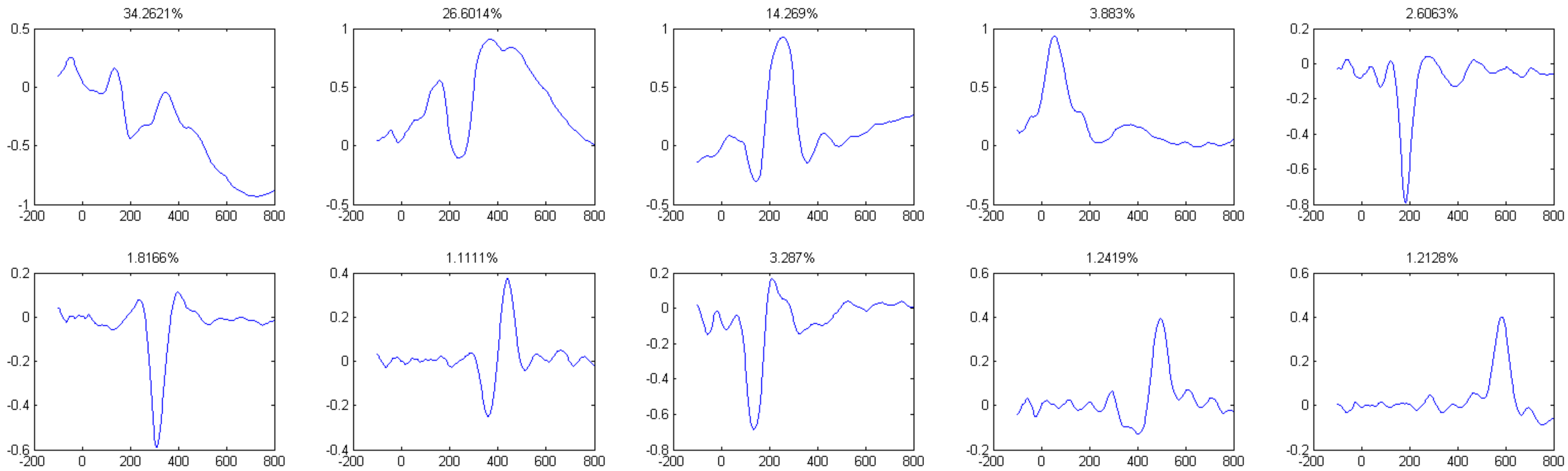
The point-by-point ANOVA for all SC-TC combinations with the factors Congruency and Numerical distance was run; very few SC-TC combinations showed significant congruency (8) or distance (5) effects. None of the SC-TC combinations was interpretable (for example, neither the second nor the third TC showed any effects of congruency or numerical distance; TFs rather resembling noise were found to be significantly modulated by experimental manipulations).



Appendix Figure VIa/1A: Spatial components and theri explained variances after varimax rotation.



Appendix Figure VIa/1B: Recomposed component scores plotted as time series for each spatial component in all conditions (virtual ERPs).



Appendix Figure VIa/1C: Temporal component loadings of the spatial component scores.

VII. The automaticity and speed of numerical processing in children[23]

Abstract

From the perspective of cognitive child development, it is an interesting question that how and when do children begin to process symbolic numerals in an automatic fashion. Armoured with an approximate magnitude representation (Brannon and Terrace, 1998; Gallistel and Gelman, 1992; Xu and Spelke, 2000; Huntley-Fenner and Cannon, 2000; Brannon et al., 2004, Xu, 2003, Wynn; 1992), young children have to learn and match arbitrary symbols to numerical magnitudes (Rubinstein and Henik, 2005; Rouselle and Noël, 2007; Ansari, 2008; Holloway and Ansari, 2009). Automaticity of processing these symbols comes with expertise (Logan, 1988; Palmeri et al., 2004), and indicates fast and obligatory access to and retrieval from memory. Expertise is built up by learning and practice – as children are getting 'experts' with symbolic numerals in school, they access and retrieve meanings of these stimuli in a fast and automatic fashion. In this study, the automaticity of number processing was tested by an involuntary paradigm: numerical information was completely irrelevant for the task. It is found that 1st grade children already process numerical information in an automatic fashion. Further, the inhibition of irrelevant information is a significant factor in young children's performance.

1. Rationale and background

1.1. Automaticity of the processing of numerical magnitudes

As introduced in chapter I (section 2.6.), one way to measure the automaticity of the access to numerical representations is the numerical Stroop paradigm. Analogous to the classical colour-word Stroop paradigm (Stroop, 1935), in the numerical Stroop paradigm (NSP) subjects have to make decisions based on the perceptual features of the

---

[23] This study is submitted (to Developmental Neuropsychology)

stimuli, while the conceptual meaning conveyed by the stimuli is not relevant. In the classical colour-word Stroop paradigm, it is the ink colour of the word what is relevant for the task. The meaning represented by the word is irrelevant and has to be ignored. Simple as it seems, the task in fact is not so easy. Word meaning still influences subjects' answers: for instance, participants slow down and make more mistakes when the word 'red' is written in green ink colour, indicating that the *semantic* and *involuntary* processing of written words is highly automatic. In the NSP, the physical sizes and numerical sizes of the two simultaneously presented stimuli are varied, and participants have to indicate the physically larger stimulus independently of their numerical meaning. Again, although a simple perceptual comparison would be sufficient to solve the task numerical meaning intrudes processing and influences one's decision; this is called the size congruity effect. The *size congruity effect* (SCE) can be delivered in the three following ways. In the *congruent* condition, numerical size difference between the two numerosities matches the physical size difference, e.g. 2 9. In the *incongruent* condition, the two dimensions (physical size and numerical size) are in conflict with each other, e.g. 2 9. In the *neutral* condition the irrelevant dimension is kept constant, e.g. 2 2 (see also Table I/1). Cost in reaction time (RT) and accuracy in the incongruent condition compared to the neutral condition is called 'interference effect'. Gain in RT or accuracy in the congruent condition compared to the neutral condition is called 'facilitation effect' (Besner and Coltheart, 1979). In adults, it is consistently found that numerical meaning is automatically processed and interferes with performance, even though it is completely irrelevant for the task (Henik and Tzelgov, 1982; Tzelgov et al., 1992; Kaufmann et al., 2005; Rubinstein and Henik, 2005; Szűcs and Soltész, 2007). The SCE reflects automatic processing of the irrelevant dimension (i.e. numerical meaning) in the given task (i.e. physical size comparison task).

Behavioural studies found that automatic access to numerical meaning builds up gradually in children. Girelli et al. (2000) found no SCE in grade 1 children, and interference effect appeared in grade 3 and got stronger in grade 5. Rubinstein et al. (2002) sampled first grade children both at the beginning and at the end of first grade in order to investigate the supposedly gradual appearance of automatic numerical processing during the first year of school. They showed that there was no SCE until the end of the first grade. Interference from the task-irrelevant numerical magnitude appeared at the end of first grade.

There are two possible explanations for the lack of SCE in the first grade. First, it might be that young children do not process numerical magnitudes automatically. Second, the absence of SCE might be due to differences in processing speed: numerical magnitudes are processed automatically, but not yet fast enough to interfere with decisions made upon physical sizes. Indeed, the relative speed of physical and numerical processing varies with age: processing of numbers becomes faster and faster with age (Noël et al., 2005). In order to test the latter, processing speed hypothesis, Mussolin and Noël (2007) balanced the processing speed differences between numerical and physical processing by letting more time for the processing of numerical information. In an NSP, they presented the two numbers at the same size for a short while. Then the numbers changed in physical size and children were asked to pick the physically larger number. The authors found that 2nd graders showed SCE; unfortunately, they did not test 1st graders. In an other study, the same authors used masked priming to control for the processing speed differences (Mussolin and Noël, 2008). They found that 2nd graders showed SCE only when the to-be-compared numbers were preceded by a prime stimulus which primed the numerical magnitudes of the target pair. Again, they did not test 1st graders. With the above very thoughtful paradigms Mussolin and Noël (2007; 2008) were able to show that numerical magnitudes are processed automatically in children, although at a slower speed than physical magnitudes. However, response time is at the end of several unseen processes and can not tell more than the presence or absence of an effect. EEG records the continuous flow of processes with a resolution in the millisecond range before (and even after) the behavioural response. So forth, EEG provides an excellent tool for exploring the exact speed and automaticity of number processing in children.

1.2. Automatic access to numerical representations: crude, or refined?

A further question besides automaticity is that in what detail numerical symbols are matched onto the magnitude representation during automatic number processing. One possibility is that only a crude categorical representation is activated (Tzelgov et al., 1992), reflecting the evaluation of the relative positiosn of the two presented numbers only on a ’smaller and bigger’ scale. An other possibility is that access to a more refined representation is carried out, and a more refined numerical relation besides being smaller or larger is evaluated. In the former scenario, if magnitudes are processed in a crude categorical manner only, the numerical distance between the given numerosities will not have an effect on decisions. According to the latter scenario, when the relations of the numerical magnitudes are evaluated in more detail, the *symbolic distance effect* will influence participants’ decisions. The symbolic distance effect reflects the analogue nature of numerical representations (Moyer and Landauer, 1967): magnitudes closer to each other are more difficult to discriminate than magnitudes further apart. Symbolic distance effect has been found when numerical information processing was voluntary. During simple number comparison (when one has to decide which number is larger or smaller) the numerical distance effect emerges in behaviour (Moyer and Landauer, 1967; Buckley and Gillman, 1974; Hinrichs et al., 1981) and in event related potentials (Dehaene, 1996; Pinel et al., 2001; Szűcs and Csépe, 2004, 2005; Soltész et al., 2007). Number comparison studies with children also showed the numerical distance effect suggesting that children process and compare numerical meaning in the similar manner to adults (Sekuler and Mierkevitz, 1977; Duncan and McFarland, 1980). 5-year-old children showed the numerical distance effect comparable to that of adults in event-related potentials (ERP) as well (Temple and Posner, 1998).

Numerical distance effect (NDE) has been found in 6-year-old children in an *involuntary* numerical task: children were asked to tell whether the two numbers shown were physically same or not (e.g. 4 4 is same, 4 2 is not the same). In this rather simple perceptual task response time was influenced by the numerical distance: it took longer to tell of a number pair close to each other that they were not the same than of a number pair further away from each other (Duncan and McFarland, 1980). However, in the NSP paradigm symbolic distance effect, so forth the activation of a refined numerical representation is not consistently found either in adults (Tzelgov et al., 1992) or in children (Rubinstein et al., 2002). The inconsistency of the distance effect in NSP can be again accounted for in two different ways. On the one hand, it is assumed that only a crude evaluation of magnitudes happens (Tzelgov et al., 1992) under certain experimental circumstances. On the other hand, access to a refined magnitude representation occurs, but the speed by which it happens may be sometimes too slow. Schwarz and Ischebeck (2003) proposed a race model according to which the task-relevant (i.e. physical) and the task-irrelevant (i.e. numerical) comparison competes until the decision is reached. When the relative speed of numerical comparison is slower, than that of the physical comparison, numerical distance might not have the time to exert an effect on behavioural responses. Again, EEG seems to be an optimal tool to explore the timing of a assumed numerical distance effect in NSP.

In a previous study (Szűcs, Soltész et al., 2007) we have already shown that grade 3 and 5 children process numerical meaning involuntary, showing NDE in event related potentials (ERPs) at around 210-230 ms after stimulus presentation.

1.3. Hypotheses of the present study

Here I set out to investigate involuntary number processing in grade 1, 2 and 3 children, co-registering electrophysiological data with behavioural measurements in an NSP paradigm where numerical information was irrelevant for the task (physical size comparison). With larger group sizes and with more focused analysis on involuntary numerical processing I expected to reveal a clearer picture of numerical development, significantly contributing to previous studies (Girelli et al., 2000; Rubinstein et al., 2002, Mussolin and Noël, 2007; Szűcs, Soltész et al., 2007). ERP serves as an excellent research tool for the real-time tracking of cognitive processes, independently of and well before the overt behavioural response. ERP can be used to disentangle perceptual and decisional processes from response organization and response conflict (see for example Temple and Posner, 1998; Szűcs, Soltész et al., 2007). To my best knowledge, there is no other study examining these age groups using both behavioural and ERP measures during *involuntary* number processing. Although number processing has been investigated in 5-year-old children in a *voluntary* paradigm (number comparison;

Temple and Posner, 1998), I consider the NSP paradigm as more adequate setting for investigating automatic number processing in children.

Size congruency effect would indicate automatic access to numerical information, while symbolic *numerical distance effect* would indicate access to a refined numerical representation instead of a simpler, categorical analysis of numbers. We can presume that growing familiarity with Arabic digits would increase speed and automaticity of access to numerical representation; although developing control processes may exert an opposite effect on automatic number processing. Older children may inhibit irrelevant information more effectively than younger children, so forth weakening the effect of numerical information on ERP and behaviour. Regarding these opposing trends in development, we expect different patterns of *interference* and *facilitation* during development. According to Posner (1978), interference reflects attentional and inhibitory processes, while facilitation indicates automatic processing. Following this framework, we expect stronger interference at younger ages, while interference weakens in favour of facilitation in older children. Regarding symbolic numerical distance effect, we expect strengthening NDE through ages as access to numerical representation becomes faster and matching symbolic numbers onto the analogue magnitude representation becomes more refined.

Finally, relation of ERP effects with behaviour was also investigated in order to reveal functional relationships among ERP measures and behaviour. I hypothesize strengthening relation of ERP symbolic distance effect and behavioural symbolic distance effect with growing age, as distance effect is getting stronger in ERPs and response organization is getting more effective in older children (Szűcs, Soltész et al., 2007).

2. Methods

Subjects. Children were recruited from primary schools in and around Cambridge, UK. All children came from middle class background and were white-Caucasian. All children had English as their first language. Written informed consent was obtained from parents and the study was approved by the institutional Cambridge Psychology Research Ethics Committee. Children received specially designed T-shirts (independently of their performance) as a reward after the experiment.

Initially, 72 children participated in the study. Prior to EEG processing, 17 children's EEG data have been discarded because they ceased to complete the experiment or made more than 45% of errors (suggesting chance performance on the task). Further 21 children's data were discarded during artefact filtering. 36 children's data entered analyses: 12 year 1 (mean age: 6.35 range: 6.04 - 6.67; 5 boys), 12 year 2 (7.35, 6.81 – 8.07; 4 boys) and 12 year 3 (8.54, 7.77 – 8.85; 8 boys) children.

2.1. Methods: Experimental paradigm and task

Procedure, stimuli and timing parameters were exactly the same as in chapter V**,** except for the task and for the neutral condition. Participants were asked to indicate the *physically* larger digit. Regarding the neutral condition, the numerical dimension has to be neutralized here: neutral pairs consisted of numerically identical, physically different digits (for example see Table I/1).

2.2. Methods: Data collection and pre-processing

Both response time (RT, in milliseconds) and accuracy data (percent correct, %) were collected and analyzed.

EEG was recorded by a 65-channel Geodesic Sensor Net. The sampling rate was 500 Hz, an online band-pass filter of 0.01-70 Hz was used. The data was band-pass filtered between 0.01-30 Hz offline, and epochs were extracted time-locked to the stimuli. Epochs were then baseline-corrected relative to the -100 to 0 ms interval before stimulus onset. Average reference was recomputed from the Cz electrode. Epochs containing voltage deviations exceeding ±150 μV relative to baseline at any of the recording electrodes were rejected. Trials contaminated by strong alpha activity (8-12

Hz) were rejected: when the power of alpha exceeded 10 standard deviations from the average alpha activity, trials were excluded from further analyses. Subjects showing strong alpha activity in comparison to the group's average were also excluded.

Trials with correct responses were kept for both ERP and behavioural analyses. For the purpose of a one-to-one matching of ERP data with behaviour, behavioural data of trials excluded from ERP analysis (during artefact rejection) were excluded from behavioural analyses.

2.3. Methods and results: Behavioural data

Both accuracy (percent correct [PC], %) and median response time (RT, millisecond) were entered into a Group[3] × Congruency[3] × Physical Distance[2] × Numerical Distance[2] mixed design ANOVA. For investigating the effect of numerical distance, a separate ANOVA without the neutral condition was also performed[24]. Greenhouse-Geisser corrected p and epsilon (ε) are reported where necessary. In order to reveal any further developmental effects, a Group[3] × Congruency[3] × Physical Distance[2] × Numerical Distance[2] repeated measures ANOVA was conducted for each Group separately when necessary. Post-hoc contrasts were calculated using the Tukey-Cramer test.

2.3.1. Accuracy

Significant main effect of Grade ($F(2,33)=4.75$, $p<0.02$) and main effect of Congruency ($F(2,66)=103.5$, $\varepsilon=0.96$, $p<0.0001$) were found. Accuracy increased with **Grade** (mean and standard error for the three Grades respectively: 73.5% (2.8), 76.1 (2.7) and 85.1% (2.8), Grade 1 committing more errors than Grade 2 and Grade 3 (post-hoc comparisons: Grade 1 vs. Grade 2: $p=0.07$ and Grade 1 vs. Grade 3: $p<0.02$) (Figure VII/1).

[24] There is only physical difference between the two stimuli in the Neutral condition; numerical distance is zero.

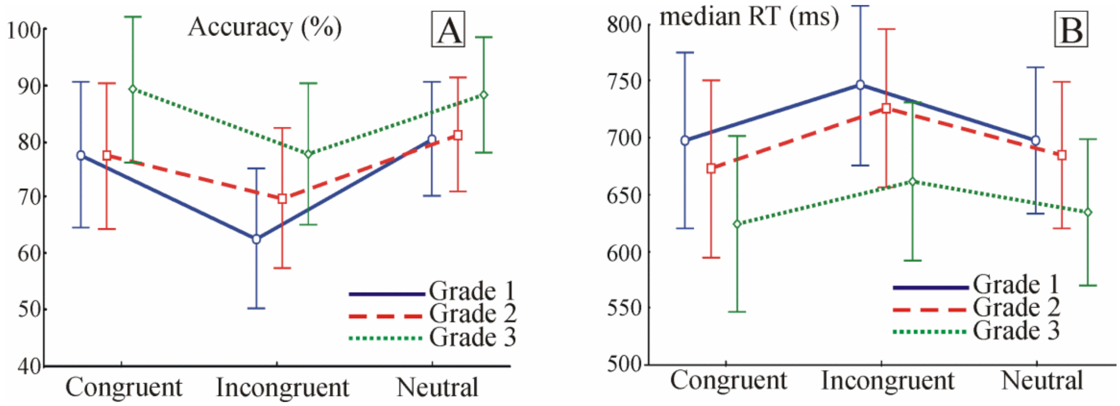
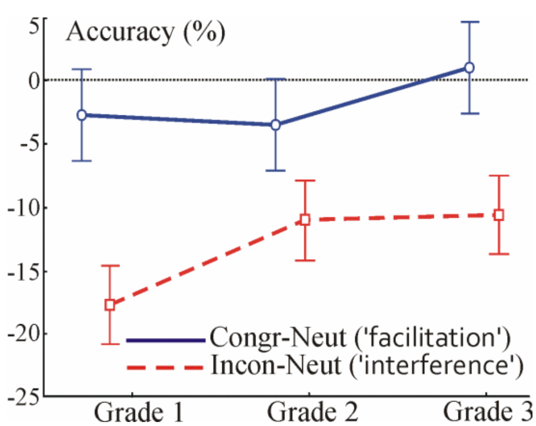


Figure VII/1: Congruency effects in accuracy and in RT.

Children performed better in the **Congruent** and in the Neutral condition than in the Incongruent condition (81.4% (3.7), 70% (3.6) and 83.3% (2.9); post hoc contrasts revealed an *interference effect*: both Congruent vs. Incongruent and Neutral vs. Incongruent: $p<0.0001$) (Fig 1). The facilitation was not significant: performance in the neutral and in the congruent condition did not differ significantly from each other ($p>0.15$).

Congruency × **Grade** interaction was significant ($F(4,66)=4.02$, $\varepsilon=0.96$ $p<0.01$, although the same pattern has been revealed in all three groups by the post hoc analyses (Figure VII/1). Incongruent condition yielded more bad responses than the Neutral and the Congruent conditions in all three groups ($p<0.0001$ for all three Grades).



The Congruency × Grade interaction is due to the stronger interference effect in Grade 1 than in Grade 2 or Grade 3. Figure VII/2 illustrates the Congruency × Grade interaction: congruent minus neutral and incongruent minus neutral difference values represent the developmental change of 'facilitation' and 'interference' across grades.

Figure VII/2: 'Facilitation' (Neutral minus Congruent) and 'interference' (Neutral minus incongruent) effects.

The interaction of **Congruency × Numerical distance** was significant (F(1,33)=20.13, p<0.0001) (Figure VII/3A).

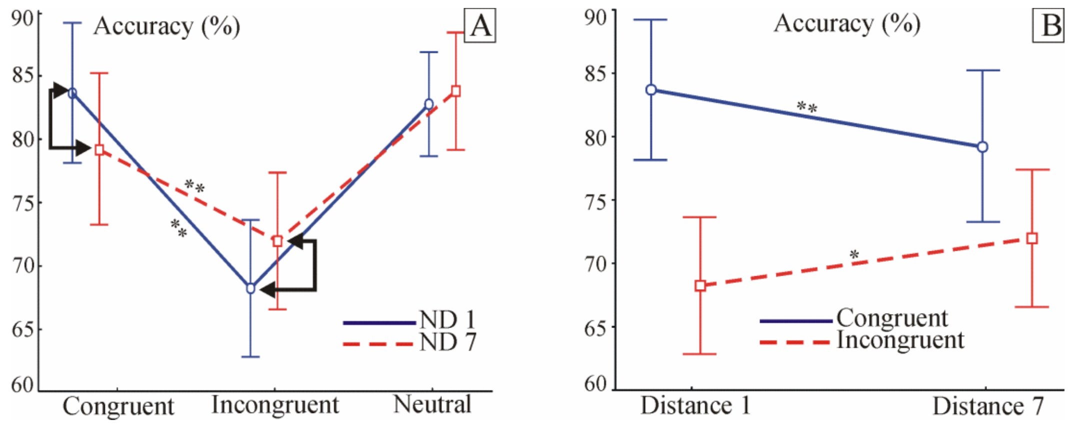


Figure VII/3: A: Congruency × Numerical distance interaction. Arrows indicate distance effects.
B: Numerical distance effects.
* denote p<0.05; ** denote p<0.01.

The effect of congruency was modulated by numerical distance in the following way: congruency effect was more enhanced in numerical distance 1 (ND1) than in numerical distance 7 (ND7), although being significant in both numerical distances (p<0.001 for both). The Congruency × Numerical distance × Grade interaction was not significant; however, as an attempt to reveal any developmental effects, the Congruency × Numerical distance interaction was tested in all three groups separately (Figure VII/4). First, the Congruency × Numerical interaction was significant in all three groups (p<0.05). Further, in Grade 1, congruency effect was significant in both numerical distances (congruent vs. incongruent: p<0.002) while the congruency effect in the large numerical distance disappeared and was significant only in the small numerical distance condition in Grade 2 and Grade 3 (ND1: p<0.001; ND7: p>0.2).

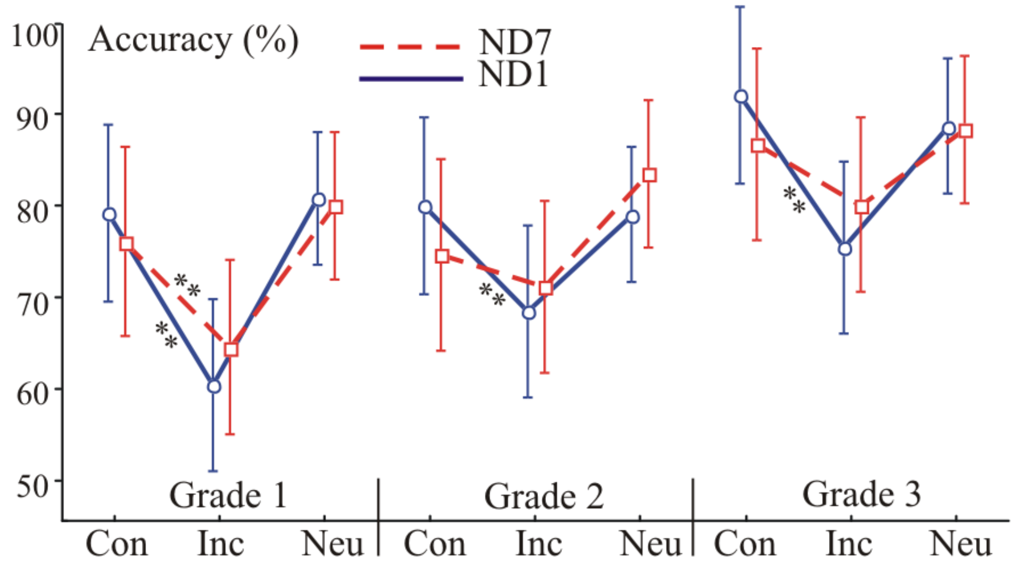


Figure VII/4: Congruency effect (congruent vs. incongruent).

For further investigating the effects of numerical distance, the neutral condition was omitted from the analysis (see methods). Post hoc comparisons showed that **numerical distance** effect was significant in both Congruent and Incongruent conditions ($p<0.01$ and $p<0.05$, respectively) (Figure VII/2B). The patterns of Numerical distance effects are consistent across age groups.

The main effect of Physical distance was also significant ($F(1,33)=106.7$, $p<0.0001$). Children committed fewer errors in the big **Physical distance** condition than in the small Physical distance condition (83.1% (3.7) and 73.4% (4.5)).

Further, a **Congruency × Physical distance** interaction was also significant ($F(2,66)=13.99$, $\varepsilon=0.95$ $p<0.0001$) (Figure VII/5), although performance in the congruent and incongruent conditions differ significantly in both physical distances ($p<0.001$ for both). The interaction is caused by the somewhat stronger effect of incongruence in the small physical distance condition.

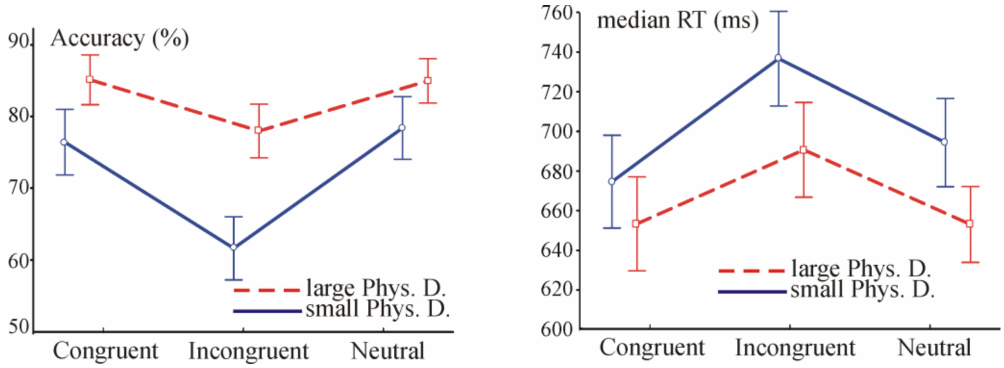


Figure VII/5: Congruency and physical distance, averaged across groups.

2.3.2. Response time

The main effects of Grade (F(2,33)=5.38, p<0.01)**,** Congruency (F(2,66)=41.71, ε=0.99, p<0.0001) and Physical distance (F(1,33)=132.4, p<0.0001) were also significant in reaction time. The main effect of **Grade** reflects response acceleration with age (711 ms (15.2), 701.6 ms (15) and 650.3 ms (15.3)), with Grade 3 being significantly faster than Grade 1 (p<0.2). The main effect of **Congruency** reflects response acceleration in the congruent and response deceleration in the incongruent condition (Congruent: 669.2 ms (18.5), incongruent: 711.6 ms (18) and neutral: 682.7 ms (17.7)). Both facilitation and interference effects were significant (Congruent vs. Neutral: p<0.002, Incongruent vs. Neutral: p<0.0001) (Figure VII/1).

The main effect of **Physical distance** reveals slower responses in the more difficult (small Physical distance: 702.7 ms (22)) condition than in the easier (big Physical distance: 672.8 ms (21.5)) condition. The effect of **Numerical Distance** was not significant either as a main effect or in an interaction in the RT data.

2.4. Methods and results: ERP data

2.4.1. Detection of ERP effects

First, trials were entered into a point-by-point Congruency[3] × Physical Distance[2] × Numerical Distance[2] repeated measures ANOVA, separately for each groups (Grade 1-3). In order to deflate the possibility of the Type I error, effects were considered as significant at alpha=0.025 across at least 8 consecutive time points (16 ms) and at least 5 electrode sites. Second, time points showing significant effect were then averaged across time, at each significant electrode and were entered into an Electrode × Congruency × Physical distance × Numerical distance ANOVA. Electrodes with negative and positive voltages were averaged separately and entered separate ANOVAs. The distinct analyses of polarities are necessary in order to avoid annulations of effects by averaging positive and negative values together. Also, in some instances, although the polarity did not change, the direction of the experimental effect showed the opposite pattern; those electrodes were also analyzed in separate blocks.

ERP analyses focus on the effect of numerical distance and the effect of congruency, as these effects are considered to reflect automatic and non-intentional number processing.

2.4.2. Automatic number processing: Congruency effects

Congruency effects were found in all three Grades (Figure VII/6). Post-hoc analyses showed that interference (incongruent vs. neutral) was more emphasized in Grade 1 children over frontal electrode sites.

Grade 1: 240-270 ms over 3 right frontal (electrode 2, 3 and 62: F(2,22)=8.53, ε=0.94, p<0.002) and 2 right centro-parietal electrode sites (43 and 47: F(2,22)=16,26, ε=0.82, p<0.001); Contrary, in Grade 2 and Grade 3 children the effect of congruency is driven by a facilitation effect (Congruent vs. Neutral) over parietal sites. Grade 2: 320-350 ms over 4 right- and centro-parietal electrode sites (46, 47, 50 and 53: F(2,22)=11.49, ε=0.94, p<0.001); Grade 3: 330-360ms at 4 right- and centro-parietal electrodes (F(2,22)=7.7, ε=0.84, p<0.004; see Figure VII/6; Figure VII/7 shows ERPs at the most typical electrodes).

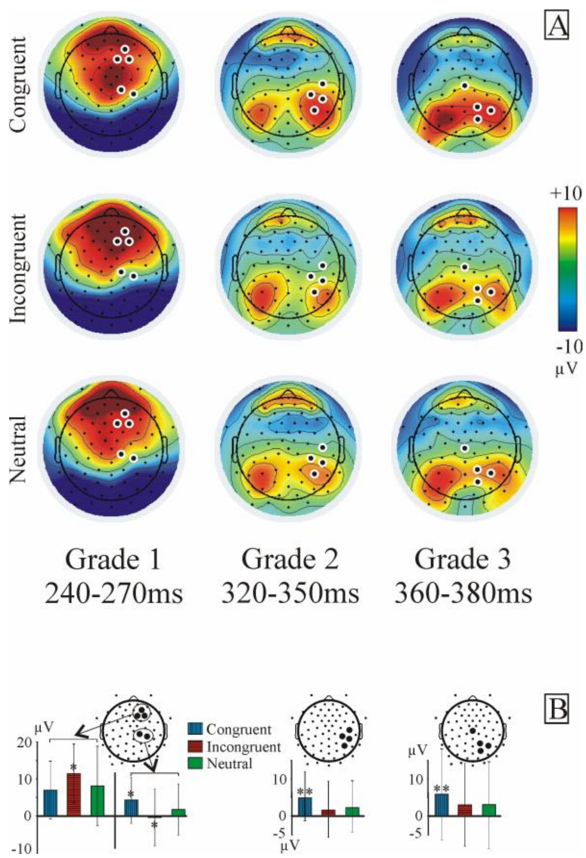


Figure VII/6: A: Topographic plots of congruency effects. Electrodes with significant effects are marked. B: Post-hoc analyses of the congruency effects. * denotes p<0.05; ** denotes p<0.01.

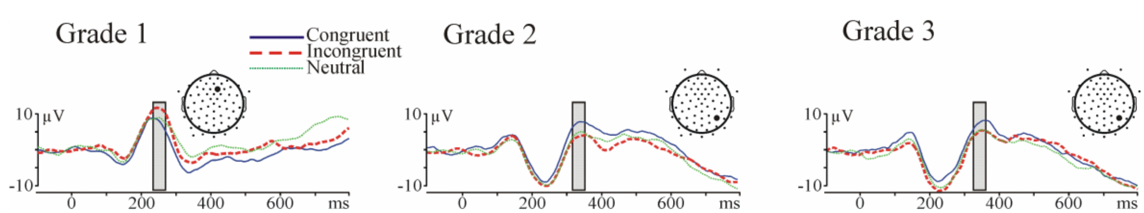


Figure VII/7: Congruency effect on the most typical electrodes. Significant time intervals are denoted.

2.4.3. The relation between ERP and behavioural interference effects

Significant effects in ERP were correlated with behavioural data in each subject in order to reveal any functional relationship between ERP effects and behaviour. For this purpose, the root mean square (RMS) of amplitudes of ERP effects were calculated and correlated with PC and RT effects. ERP effects of congruency were generated by the following way: neutral condition was subtracted both from the incongruent condition ('interference') and from the congruent condition ('facilitation'). ERP effects of numerical distance were calculated in a similar way: distance 1 was subtracted from distance 7. RMS was calculated for the purpose of eliminating the polarity and direction of ERP effects at different electrode sites. The absolute size of ERP effects were extracted across all significant electrodes, independently of their polarity allowing for an inclusive analysis of ERP effects.

In Grade 1 children the effect size of ERP 'interference' (incongruent – neutral) correlated with the effect size of 'interference' in RT ($r=-0.5847$, $p<0.05$) (Figure VII/8). The larger the ERP 'interference' effect in a child, the smaller the cost of incongruence in his RT. The correlation of ERP and behavioural interference weakened with age ($r=-0.56$, $p>0.05$ for Group 2 and $r=-0.31$, $p>0.3$ in Group 3).

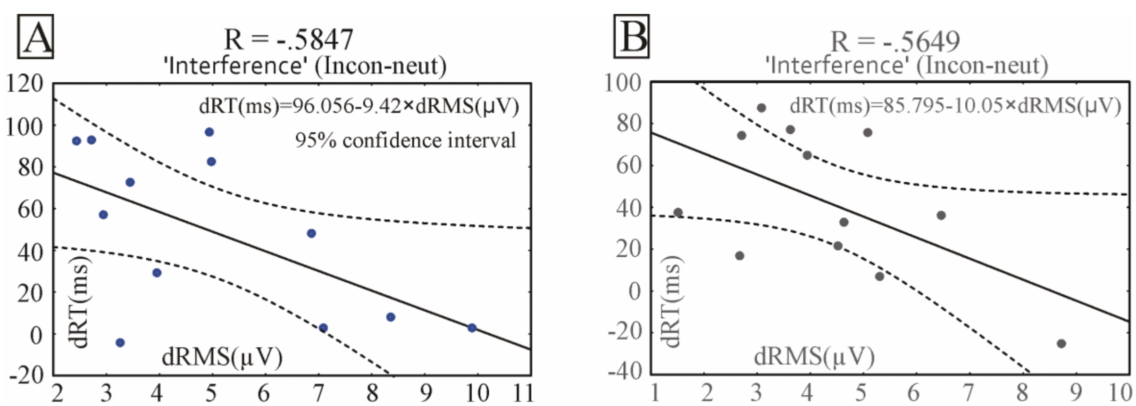


Figure VII/8: Correlation of 'interference' effects between ERPs and behaviour in Grade 1 (A) and in Grade 2 (B) (n.s.). X-axis: difference of RMS between incongruent and neutral conditions. Y-axis: difference of RT between incongruent and neutral conditions.

2.4.4. Automaticity and representation: Congruency × Numerical distance interaction

The effect of congruency was modulated by numerical distance in all three grades (Figure VII/9A). Grade 1: 220-250 ms, at 4 electrodes (34, 41 and 42: $F(1,11)=9.35$, $p<0.02$; 24: $F(1,11)=5.98$, $p<0.04$). Grade 2: 230-260 ms, at 7 electrodes (38, 40 and 45: $F(1,11)=8.26$, $p<0.02$; 3, 16, 17 and 22: $F(1,11)=11.22$, $p<0.01$). Grade 3: 190-290, at 7 electrodes (37, 38, 41, 42, 29 and 34: $F(1,11)=11.46$, $p<0.01$; 3: $F(1,11)=7.85$, $p<0.02$). Post-hoc analyses revealed significant numerical distance effects in Grade 2 and Grade 3 in the incongruent condition; numerical distance effect was not significant in Grade 1 (see Figure VII/9B). Significant distance effect arose from stronger congruency effects in large numerical distance compared to congruency effects in small numerical distance.

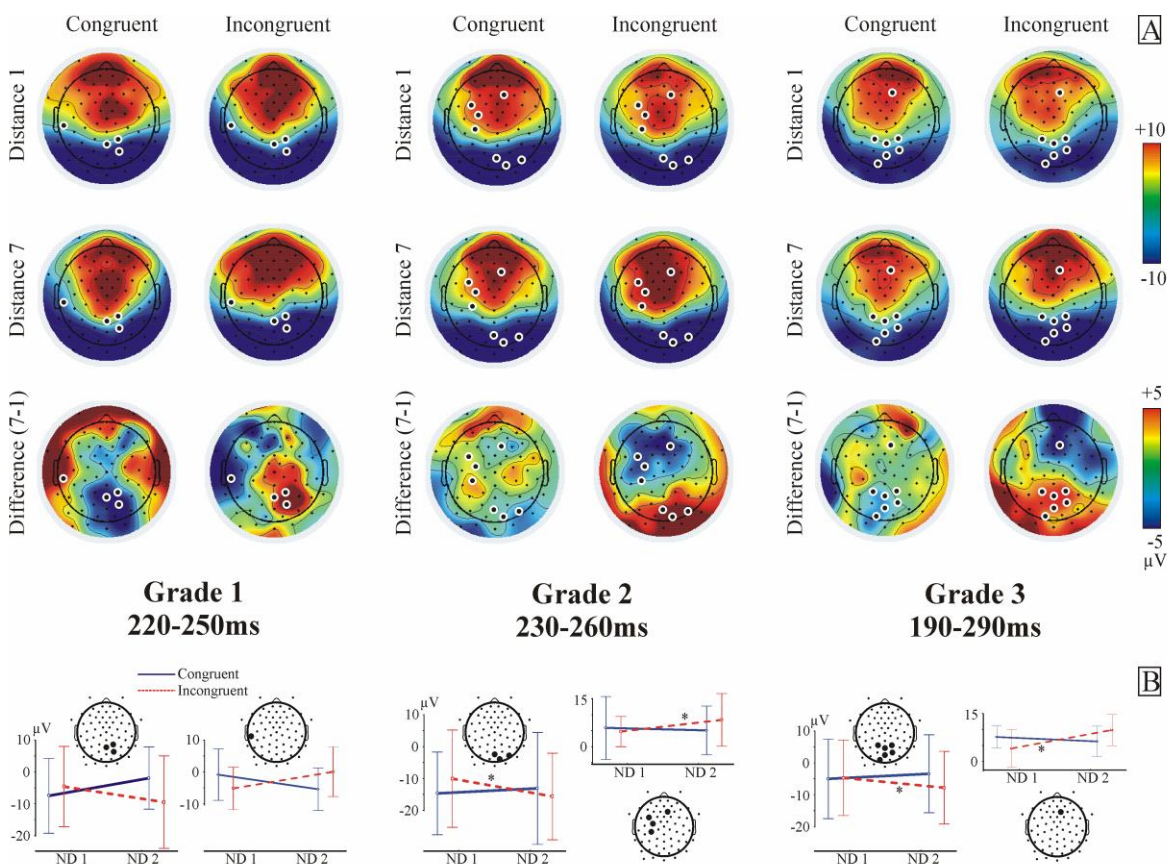


Figure VII/9: A: topographic plots of the congruency × numerical distance interaction. Significant electrodes are marked.
B: Post-hoc analyses of the congruency × numerical distance interactions.

Figure VII/10A illustrates the interaction in global field power (GFP): GFP visualizes congruency effects in a compact way. GFP is computed as the mean potential deviation of all recording electrodes, and it reflects the spatial standard deviation of the data (Lehmann and Skrandies, 1980; Skrandies, 1995). ERPs (on all channels) with high peaks and troughs and steep potential gradients are accompanied by high GFP, while flat ERPs, or minor effects present only at a few electrode sites are associated with small GFP. Therefore the GFP is an excellent tool for summarizing large and robust ERP effects at many electrodes in the form of a single curve. Also, Figure VII/10B shows the congruency × numerical distance interaction at the most typical electrodes.

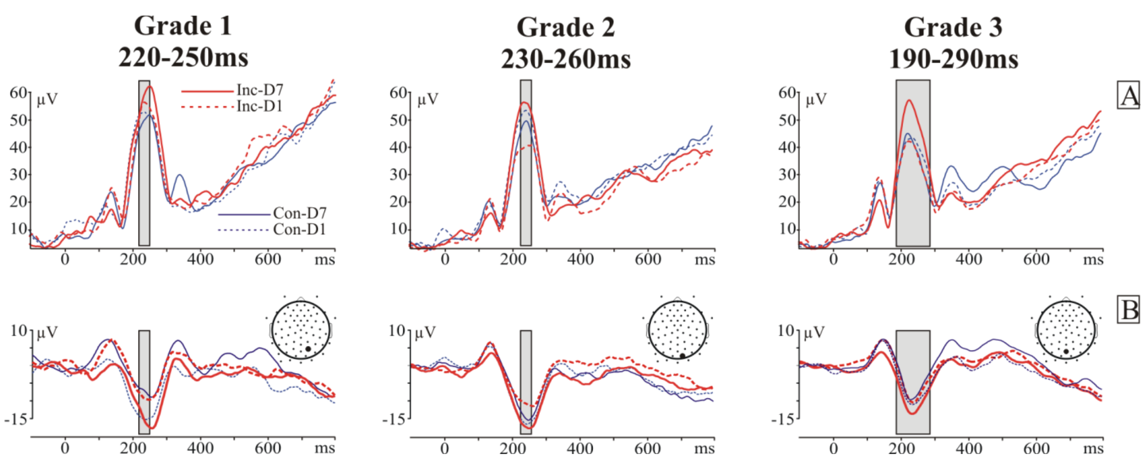


Figure VII/10: A: the congruency × numerical distance interaction on GFP.
B: congruency × numerical distance interaction on the most typical electrodes.

2.4.5. The relation between ERP and behavioural numerical distance effects

In Grade 3, the size of the ERP distance effect in the congruent condition (in RMS; see methods) correlated with the size of the RT numerical distance effect in the congruent condition ($r=-0.599$, $p<0.04$) (Figure VII/11). No such interaction emerged in the younger groups.

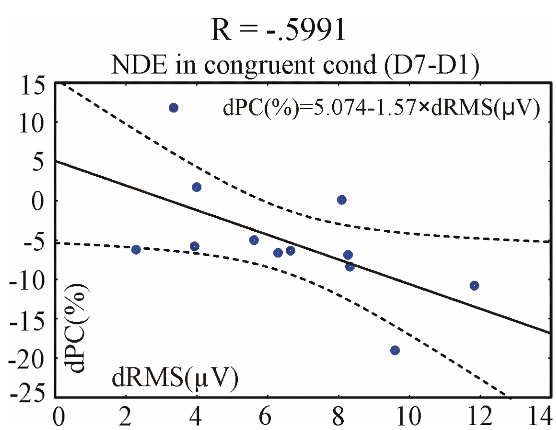


Figure VII/11: Correlation of the size of numerical distance effect (NDE) in accuracy and in ERP.

3. Discussion of the results

3.1. Automatic number processing

The significant *size congruity effect*s (SCE) both in behaviour and in ERPs confirm that children access numerical representations *involuntarily*. In behaviour it was found that irrelevant numerical meaning intrudes physical decisions already in Grade 1 children (6 years), suggesting that symbolic Arabic numerals are highly familiar and are processed involuntary already at age 6. Furthermore, the effect of *interference* is the strongest in the youngest group (Grade 1) and weakens with age (in Grade 2 and 3). ERP data confirms and adds to behavioural findings. In Grade 1, there is a strong interference effect at around 240-270 ms, over frontal electrode sites, probably suggesting the utilization of attentional and inhibitory processes. Meanwhile, facilitation effect dominated the SCE in grade 2 and in Grade 3 over the right parietal electrodes in the range of the P3 component. These ERP results suggest that Grade 2 and 3 children were more successful in inhibiting irrelevant information when it was in conflict with the task than Grade 1 children – although interference was still present in their behavioural response.

The correlation between ERP interference effect and RT interference effect in Grade 1 children further suggest the importance of inhibitory and attentional functions at this young age. More precisely, the larger was the interference effect in ERPs, the smaller was the cost of interference in RT. This finding can be interpreted in the following way. The ERP effects around 240-270 ms reflects the involvement of functions related to attention or inhibition. Children who showed stronger effects in ERPs, managed to inhibit the interference of irrelevant information somewhat better than children who did not show such a strong effect in ERPs.

In summary, numerical meaning is processed *involuntarily* and in an automatic fashion already in grade 1 children. In comparison to Grade 1, the interference effect weakens in Grade 2 and 3 in behaviour. Also, ERP congruency effects are dominated by facilitation and not interference. Referring to Posner's interpretation of facilitation and interference, probably *facilitation* and *interference* reflects two different sources of cognitive development. Attentional and inhibitional processes get more emphasis at younger ages, due to the relatively immature functioning if the prefrontal cortex (Bunge et al., 2002; Durston et al., 2002; Adleman et al., 2002; Schroeter et al., 2004). Meanwhile, numerical processing becomes more automatic during development and emphasis shifts from inhibition towards automatic processing.

3.2. Numerical representation: crude or refined?

Numerical distance effect (NDE) was not found to be significant in behavioural data. This is in concert with previous developmental studies not finding NDE during *involuntary* number processing in young children (Girelli et al., 2000; Rubinsten et L., 2002; Szűcs, Soltész et al., 2007). This would suggest a crude access to numerical representations, discriminating magnitudes on a categorical basis (Tzelgov et al., 1992). In the Szűcs, Soltész et al (2007) study we did not found significant NDE in ERPs neither; however, only the main effect was considered. In the present study not only the main effects, but the interactions of numerical distance effect were also tested. Irrelevant numerical distance modulated congruency effects both in behaviour and in ERPs; the congruency × numerical distance interaction was significant in all three grades. The timing of this interaction in ERPs is in line with earlier studies, showing NDE at around 200 ms after stimulus presentation (e.g. Temple and Posner, 1998; Grune et al., 1993; Dehaene et al, 1998). So forth irrelevant numerical distance interacted with the physical decision of children, already at grade 1. This suggests that a refined representation of numerical magnitudes is accessed in children.

The NDE in ERPs correlated with the NDE in accuracy in grade 3 children: the larger the difference between the two numerical distances in ERP, the larger the difference in the NDE in accuracy. The correlation was significant only in the congruent condition – it would be speculative to draw firm conclusions. Probably the recognition of a congruent, so forth facilitatory relation between physical size and numerical magnitude, signalled by the ERP effect, led to a stronger behavioural gain in these congruent situations. This is in line with he fact that ERP NDE and behavioural NDE correlated significantly only in the third grade suggest the growing automaticity of refined numerical processing with age.

4. Conclusion

It has been shown that 1st grade children already process numerical magnitudes automatically, reflected by the influence of numerical distance to their overt decisions and ERPs in a physical comparison task. If physical properties had been more salient than numerical information to children, then the interference from the irrelevant numerical information would have been smaller in grade 1 than for example in grade 3. But this is clearly not the case; interference was in fact the strongest in 1st grade and weakened with age. This suggests that numerical information is as salient as physical properties to children.

Correlations among ERP measures and behaviour showed that the amount of interference from the task-irrelevant numerical information in ERPs predicted the cost of interference from numerical information in response time. This correlation gradually weakened across the three grades and indicates the significance of attention and inhibition in younger children, even in a seemingly simple task like physical size comparison.

Parallel to the development of the inhibition of interfering irrelevant information, the automatic processing of numerical information get somewhat faster in the 3rd grade. Also, numerical distance effect predicted behavioural distance effect in grade 3 children, suggesting a strong and predictive link between the ERP and behavioural markers of numerical processing.

Conclusions and future directions

1. Modularity and magnitude representation

As it has been discussed in chapter I, it is a commonly held view that the representation of magnitudes is a neuro-cognitive module (e.g. Dehaene and Cohen, 1995, 1997); it is *innate*, *domain specific*, *hardwired*, *informationally encapsulated* and the access to it is *fast* and *automatic*. It is supposed to be innate, as several studies with animals and human infants have concluded. Both animals and non-verbal infants are able to discriminate approximate magnitudes based on their numerosity, moreover, their discrimination abilities obey the same regularities (the ratio effect and Weber's law; e.g. Xu and Spelke, 2000; Cantlon and Brannon, 2007). The horizontal part of the intraparietal sulcus (IPS) has been found to be activated during magnitude- and number tasks, in several experiments – so forth the IPS is the proposed host for the representation of numerical magnitudes in the brain (Dehaene et al., 2003). Number representation was also found to be isolated from other representations: adults who suffered brain damage to the parietal cortex showed deficiencies in their numerical abilities, while their other high-level cognitive functions remained intact (Dehaene and Cohen, 1995). Fast and automatic access to the numerical representation has also been proved – the effect of numerical distance exerts its effect already at around 200 ms both in children and in adults (ERP studies; Grune et al., 1993; Dehaene, 1996; Szűcs, Soltész et al, 2007; chapter V and chapter VI).

However, there are of course studies which rather confute the modular nature of numerical cognition (see chapter I). In sum, after the review of the literature I would suggest to conclude that the supposedly modular magnitude representation is not straight away hand in hand with numerical abilities; there is rather a large and complex network of different abilities, including the analogue magnitude representation as well, which support the acquisition of counting and mathematics. Regarding magnitude representation, it seems to be more like a general tool or measure for all kinds of perceptual and analogue magnitudes which are covariates of numbers among natural circumstances (e.g. sum surface area). This basic ability to discriminate continuous magnitudes may serve as a departure point for the understanding of concepts like 'more' and 'less', 'add' and 'take away', which are part of later arithmetics.

2. Magnitude representation in developmental dyscalculia

As it has been discussed in chapter III, there is no consensus on the causal factors of developmental dyscalculia (DD). The *defective number module hypothesis* assumes that the core 'number module' is impaired in DD (e.g. Butterworth, 1999). Contrary, some emphasize the importance of *more domain-general abilities*, like working memory and attention deficits in DD. In the two studies investigating DD (Chapter III) it was found that the marker of the *analogue magnitude representation,* the *numerical distance effect,* was indeed intact in DD – suggesting that it is probably not the core magnitude representation which is deficient in DD adolescents. Detailed ERP and principal component analysis showed that the numerical distance effect was similar in pattern and in timing in DD adolescents and in their controls. Differences between the two groups arose in some principal components and also at the later stages of processing measured by ERPs, suggesting the involvement of monitoring and executive functioning. These assumptions are in line with the behavioural results: DD adolescents were although just as accurate as their peers in a task measuring executive functions, they were significantly slower; suggesting processing and/or strategic differences when executive functions are taxed. As their accuracy was not lower than the control group's, these differences in executive abilities may not be so apparent in other fields of performance. As the *neuroconstructivist perspective* also suggest, slight deficits of some basic abilities may not be recognizable in everyday life, but can lead to apparent difficulties in more complex functions. And mathematics is clearly a highly complex field; attention, working memory, executive functions are all intensively taxed during performing arithmetic tasks. In the future, I would put much more emphasize on the detailed investigation of these basic abilities in DD.

Besides, a slight impairment of the mental representation of fingers was also found, suggesting some probably more specific deficits in DD as well. It may be that a specific *network* involving both basic functions (like attention and executive functioning) and specific abilities, like finger representation (and probably other spatial functions as well – all linked to the parietal lobes, e.g. Gerstmann syndrome), is slightly impaired and led to DD. The exact nature of the interactions and causal routes of these deficits and weaknesses and how they lead to difficulties in learning of mathematics would clearly require the *longitudinal investigation* of these abilities both in normal and in DD children.

3. The development of magnitude representation and counting in children

Chapter IV investigated the co-development of the analogue magnitude representation, counting, verbal abilities and memory in kindergarten children. The study was inspired by the debate whether *counting abilities* arose straight from the analogue magnitude representation (e.g. Halberda et al., 2008), or *verbal abilities* were needed for the acquisition of the concept of discrete numbers and counting (e.g. Rouselle et al., 2004).

After a critical review of the literature on the numerical abilities of infants (including the milestone study of Xu and Spelke, 2000), I attempted to design an experimental paradigm in which the confounds arising from the unavoidable correlations among perceptual properties and numerical magnitudes are better controlled for. This control is important in order to separate children's perceptual discrimination abilities from their numerical discrimination abilities. It was found that verbal counting abilities (i.e. knowing how to count) correlated with performance in the approximate magnitude comparison task, but only at the very beginning of the acquisition of these verbal abilities. This led to the conclusion that formal arithmetic education, and verbal counting knowledge in peculiar helps children to understand that number is an abstract property of objects and can be quantified by discrete, exact values. This knowledge may further help them to pay attention to numbers independent of other physical and perceptual properties.

But of course, alternatively one could say that it is the other way around: the refinement of the approximate magnitude representation helps children to learn how to count and later arithmetics (for example Halberda et al., 2008). However, if it were so, we should also conclude that native Americans do not learn to count and to use number words because their innate abilities do not allow them to do so: as it has been shown, their approximate representation is less refined than that of Western adults'. This argument is at least awkward. The relation between magnitude representation and counting, if causal, is rather that the verbal tools acquired via formal education lead to a more polished representation of approximate magnitudes.

The importance of general abilities, like working memory was found to be relevant both in typical and atypical numerical development (chapters III and IV), suggesting that some other abilities than the representation of approximate magnitudes play significant roles in the acquisition of arithmetics.

4. The interaction of physical size and numerical magnitude in children and in adults

In chapter V, the development and interactions of magnitude and number representation were examined in a numerical Stroop paradigm, where task-irrelevant physical size and task-relevant numerical size was manipulated in an orthogonal fashion. Behavioural and ERP results confirmed that grade 1-3 children represent numerical magnitudes in a similar manner to adults; *numerical distance* exerted its effects at around 170 ms after stimulus presentation in adults and at around 200 ms in children. This is in line with previous studies (e.g. Temple and Posner, 1998; Szűcs, Soltész et al., 2007) and suggests a *fast access* to the representation of numerical magnitudes. The slower and more error-prone responses of children are due to the stronger susceptibility to *interference* from irrelevant information, arising from the weaker executive and inhibition abilities. Stronger interference effects in children were found by both behavioural and ERP measures.

Furthermore, as an explorative methodological extension for this study, a spatiotemporal principal component analysis was performed on adults' data. PCA disentangled at least two separate *latent components* behind facilitation and interference effects in adults. This is an interesting and promising finding with regard to the complexity of Stroop tasks.

The question whether numbers and physical size overlap at the *representational* or at the *decisional level,* unfortunately cannot be answered with this paradigm. Even if the processing of either of the dimensions (i.e. the physical size in this case) was involuntary, all we can say is that the processing of this information happened in an *involuntary* fashion. Probably a paradigm where comparison is not necessary at all could help to answer whether a common mental measure or a common mental representation leads to the interaction. Such a paradigm could be an *adaptation* paradigm, where decisions on magnitudes are not required to make, while brain responses to changes in magnitude can still be measured. Some studies with the adaptation method have already been undertaken (Piazza et al, 2004, 2007; Cantlon et al., 2006); however, there were some severe confounds in these studies (for discussion, see chapter I). A study using adaptation is actually under its way; I attempted to control more perceptual features than previous studies did, and manipulated the task requirements in order to divert participants' attention as far as possible from numerical magnitudes.

5. The automaticity of number processing in children

The study presented in chapter VI was carried out in order to investigate the *speed* and *automaticity* of the access to numerical representations in young children. This is an interesting issue because a fast and accurate link between symbolic numerals (Arabic digits in the present case) and their referents, numerical magnitudes is inevitable for the learning of arithmetics. The efficiency of the matching of these symbols with their meanings in memory was found to be predictive for school performance during the earliest years of school (Holloway and Ansari, 2009) and was also found to be relevant in developmental dyscalculia (Rouselle and Noël, 2007).

Using the physical version of the numerical Stroop paradigm, it was found that grade 1 children already process the meaning of and relations between Arabic numerals *involuntarily*. Furthermore, numerical meaning not only interfered with, but also modulated physical judgements in function of the symbolic *numerical distance.* This finding leads to the conclusion that not only a crude evaluation of the relation of numerical magnitudes, but a refined comparison occurs in a fast and automatic fashion in children.

Interference from the irrelevant numerical information was in fact the strongest in the youngest group (grade 1) and weakened by age. This suggests that the *automatic* processing of numerical symbols is very fast and unavoidable and that, even being a little counterintuitive (and also contradicts to Piaget's theory), numerical magnitudes are just as salient for young children as physical/perceptual magnitudes are. The strong interference effects in grade 1 children also suggest that *inhibitory functions* are weaker in younger children and develop year by year.

In parallel with the decreasing interference across grades, the size of the ERP interference effect correlated with the size of the interference effect in behaviour in grade 1. This correlation weakened and lost its significance with age. This finding may also reflect the emphasized role of attention and inhibitory functions at younger ages. Besides the development of inhibitory abilities, automatic number processing got also faster with age. In accord with this growing automaticity, the size of the numerical distance effect in ERP predicted behaviour.

As for a technical note at the end, measuring the size of experimental effects in ERP was proven to be promising for the linking of ERP signatures and behaviour; in the future, ERP measurements may be able to track and *predict* the development of cognitive abilities in children.

6. Conclusion

I try to keep this very last conclusion as short as possible; there have already been several conclusions throughout the thesis.

To sum up, I think it is oversimplifying to attribute all the relevance to the *analogue magnitude representation* in the (typical and atypical) development of numerical cognition. In fact, many researchers in the field of cognitive neuroscience disregard the importance of many other abilities necessary for numerical development. *Cognitive neuroscience* and *educational research* have to be moved closer to each other in this matter. Numerical development and the learning of mathematics rely on a broad *network* of cognitive functions: especially working memory, attention, and inhibitory functions seem to be very important and these abilities were found to play a possible role in atypical development as well (developmental dyscalculia). And, although not in the scope of the present thesis, other than purely cognitive factors are also very important (emotions, self-esteem, motivation, interest, social background etc).

Let me finish with an immensely amusing citation of Albert Einstein:

- "*Do not worry about your difficulties in mathematics; I assure you that mine are greater.*" -

Related, first-authored publications

Soltész, F., & Szűcs, D. (2009). An electro-physiological temporal principal component analysis of processing stages of number comparison in developmental dyscalculia. Cognitive Development, 24(4), 473-485.

Soltész, F., Szűcs, D., Dékány, J., Márkus, A., & Csépe, V. (2007). A combined event-related potential and neuropsychological investigation of developmental dyscalculia. Neuroscience Letters, 417(2), 181-6.

Soltész, F., Szűcs, D., & Szűcs, L. (revised, re-submitted). Relationships between magnitude representation, counting and memory in 4- to 7-year-old children: A developmental study. Behavioral and Brain Functions.

Soltész, F., Szűcs, D., White, S. (submitted). Event-related brain potentials dissociate the developmental time-course of automatic numerical magnitude analysis and cognitive control functions during the first three years of primary school. Developmental Neuropsychology.

Soltész, F., Goswami, U., White, S., Szűcs, D. (submitted). Executive function effects on numerical development in children: Behavioural and ERP evidence from a numerical Stroop paradigm. Learning and Individual Differences.

Related, second-authored publications

Szűcs, D., Soltész, F., & White, S. (2009). Motor conflict in Stroop tasks: direct evidence from single-trial electro-myography and electro-encephalography. Neuroimage, 47(4), 1960-73.

Szűcs, D., Soltész, F., Bryce, D., & Whitebread, D. (2009). Real-time tracking of motor response activation and response competition in a Stroop task in young children:

a lateralized readiness potential study. Journal of Cognitive Neuroscience, 21(11), 2195-206.

Szűcs, D., & Soltész, F. (2008). The interaction of task-relevant and task-irrelevant stimulus features in the number/size congruency paradigm: an ERP study. Brain Research, 143-58.

Szűcs, D., & Soltész, F. (2007). Event-related potentials dissociate facilitation and interference effects in the numerical Stroop paradigm. Neuropsychologia, 45(14), 3190-202.

Szűcs, D., Soltész, F., Jármi, E., & Csépe, V. (2007). The speed of magnitude processing and executive functions in controlled and automatic number comparison in children: an electro-encephalography study. Behavioral and Brain Functions, 30, 3-23.

References

Adleman, N., Menon, V. E., Blasey, C. M., White, C. D., Warsofsky, I., Glover, G. H., et al. (2002). A developmental fMRI study of the Stroop color-word task. Neuroimage, 16, 61-75.

Alarcón, M., DeFries, J.C., Light, J.G., & Pennington, B.F. (1997). A twin study of mathematics disability. Journal of Learning Disabilities, 30, 617-623.

Andreassi, J.L. (2006). Psychophysiology: Human behaviour and psychological response. Psychology Press.

Ansari, D. (2008). Effects of development and enculturation on number representation in the brain. Nat Rev Neurosci, 9(4), 278-291.

Ansari, D. & Dhital, B. (2006). Age-related changes in the activation of the intraparietal sulcus during non-symbolic magnitude processing: an event-related fMRI study. Journal of Cognitive Neuroscience, 18, 1820-28.

Ansari, D., & KArmiloff-Smith, A. (2002). Atypical trajectories of number development: a neuroconstructivist perspective. Trends in Cognitive Sciences, 6(12), 511-516.

Ansari, D., Dhital, B., & Siong, S.C. (2006a). Parametric effects of numerical distance on the intraparietal sulcus during passive viewing of rapid numerosity changes. Brain Research, 1067(1), 181-8.

Ansari, D., Fugelsang, J.A., Dhital, B., & Venkatraman, V. (2006b). Dissociating response conflict from numerical magnitude processing in the brain: an event-related fMRI study. NeuroImage,32(2), 799-805.

Ansari, D., Garcia, N., Lucas, E., Hamon, K., & Dhital, B. (2005). Neural correlates of symbolic number processing in children and adults. Neuroreport, 16, 1769-1773.

Antell, S. A., & Keating, D. P. (1983). Perception of numerical invariance in neonates. Child Development, 54, 695-701.

Arbuthnott, K., & Frank, J. (2000). Trail Making Test, Part B as a measure of executive control: Validation using a set-switching paradigm. J. Clin. Exp. Psychopathol., 22, 518-528.

Ashkenazi, S., Henik, A., Ifergane, G., & Shelef, I. (2008). Basic numerical processing in left intraparietal sulcus (IPS) acalculia. Cortex, 44(4), 439-448.

Ashkenazi, S., Rubinsten, O., & Henik, A. (2009). Attention, Automaticity, and Developmental Dyscalculia. Neuropsychology, 23(4), 535-540.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), The psychology of learning and motivation: Advances in research and theory (Vol. 8, pp. 47-89). New York: Academic Press.

Banks, W.P., & Flora, J. (1977). Semantic and perceptual processes in symbolic comparisons. Journal of Experimental Psychology: Human Perception and Performance, 3(2), 278-90.

Barrouillet, P., Fayol, M., & Lathulie`re, E., 1997. Selecting between competitors in multiplication tasks: An explanation of the errors produced by adolescents with learning difficulties. International Journal of Behavioral Development 21, 253–275.

Barth, H., Kanwisher, N., & Spelke, E. (2003). The construction of large number representation in adults. Cognition, 86, 201-221.

Barth, H., La Mont, K., Lipton, J., & Spelke, E. S. (2005a). Abstract number and arithmetic in preschool children. Proceedings of the National Academy of Sciences of the United States of America, 102(39), 14116-14121.

Barth, H., La Mont, K., Lipton, J., Dehaene, S., Kanwisher, N., & Spelke, E. (2005b). Non-symbolic arithmetic in adults and young children. Cognition, 98, 199-222.

Berger, H. (1929). Über das Elektroenkephalogramm des Menschen. Archiv für Psychiatrie und Nervenkrankheiten, 87: 527-570.

Besner, D., & Colthart, M. (1979). Ideographic and alphabetic processing in skilled reading of English. Neuropsychologia, 17, 467-472.

Bijeljac-Babic, R., Bertoncini, J., & Mehler, J. (1993). How do four-day-old infants categorize multisyllabic utterances? Developmental Psychology, 29, 711-721.

Booth, J. L., & Siegler, R. S. (2006). Developmental and Individual Differences in Pure Numerical Estimation. Developmental Psychology, 42(1), 189-201.

Booth, J.L., & Siegler, R.S. (2008). Numerical magnitude representations influence arithmetic learning. Child Development, 79(4), 1016-31.

Boxtel, G. J. M. (1998). Computational and statistical methods for analyzing event-related potential data. Behavior Research Methods, Instruments, & Computers, 30, 87-102.

Brannon, E. M. (2003). Number knows no bounds. Trends Cogn Sci, 7(7), 279-281.

Brannon, E. M., & Terrace, H. S. (1998). Ordering of the numerosities 1 to 9 by monkeys. Science, 23, 746-749.

Brannon, E. M., & Van de Walle, G. A. (2001). The development of ordinal numerical competence in young children. Cogn Psychol, 43(1), 53-81.

Brannon, E. M., Abbot, S., Lutz, D. J. (2004). Number bias for the discrimination of large visual sets in infancy. Cognition, 93, B59-B68.

Brannon, E. M., Lutz, D., & Cordes, S. (2006). The development of area discrimination and its implications for number representation in infancy. Dev Sci, 9(6), F59-64.

Brannon, E.M., Libertus, M.E., Meck, W.H., & Woldorff, M.G. (2008). Electrophysiological measures of time processing in infant and adult brains: Weber'sLaw holds. Journal of Cognitive Neuroscience, 20(2), 193-203.

Buckley, P. B., & Gillman, C. B. (1974). Comparison of digits and dot patterns. Journal of Experimental Psychology, 103, 1131-1136.

Bull, R., Espy, K.A., & Wiebe, S.A. (2008). Short-term memory, working memory, and executive functioning in preschoolers: Longitudinal predictors of mathematical achievement at age 7 years. Developmental Neuropsychology, 33(3), 205-228.

Bull, R., Johnston, R. S., & Roy, J. A. (1999). Exploring the roles of the visual-spatial sketch pad and central executive in children's arithmetical skills: Views from cognition and developmental neuropsychology. Developmental Neuropsychology, 15, 421–442.

Bull, R., Phillips, L. H., & Conway, C. A. (2008). The role of control functions in mentalizing: Dual-task studies of Theory of Mind and executive function. Cognition, 107(2), 663-672.

Bunge, S. A., Dudukovic, N. M., Thomason, M. E., Vaidya, C. J., & Gabrieli, J. D. E. (2002). Immature frontal lobe contributions to cognitive control in children: evidence from fMRI. Neuron, 33, 301-311.

Burgio-Murphy, A., Klorman, R., Shaywitz, S. E., Fletcher, J. M., Marchione, K. E., Holahan, J., et al. (2007). Error-related event-related potentials in children with attention-deficit hyperactivity disorder, oppositional defiant disorder, reading disorder, and match disorder. Biological Psychology, 75, 75-86.

Butterworth, B. (1999). The Mathematical Brain, Macmillan, London.

Calabria, M., & Rossetti, Y. (2005). Interference between number processing and line bisection: a methodology. Neuropsychologia, 43(5): 779-83.

Campbell, J.I.D., & Clark, J.M. (1988). An encoding complex view of cognitive number processing: Comment on McCloskey, Sokol & Goodman (1986). Journal of Experimental Psychology: General, 117, 204-214.

Cantlon, J. F., Brannon, E. M., Carter, E. J., & Pelphrey, K. A. (2006). Functional imaging of numerical processing in adults and 4-year-old children. PLOS Biology, 4, 844-854.

Cantlon, J. F., Platt, M. L., & Brannon, E. M. (2008). Beyond the number domain. Trends in Cognitive Sciences,.

Cantlon, J.F., & Brannon, E.M. (2007). Basic math in monkeys and college students. PloS Biology, 5(12), e328.

Cantlon, J.F., Brannon, E.M., Carter, E.J., & Pelphrey, K.A. (2006). Functional imaging of numericl processing in adults and 4-y-old children. PLoS Biology, e125.

Cappelletti, M., Freeman, E. D., & Cipolotti, L. (2009). Dissociations and interactions between time, numerosity and space processing. Neuropsychologia, 47(13), 2732-2748.

Carey, S., & Xu, F. (2001). Infants' knowledge of objects: beyond object files and object tracking. Cognition, 80(1-2), 179-213.

Carr, T.H. (1992). Automaticity and cognitive anatomy: is word recognition "automatic"? American Journal of Psychology, 105(2), 201-307.

Cattell, R. B. (1966). The sree test for the number of factors. Multivariate Behavioral Research, 1, 245-276.

Chapman, R. M., & McCrary, J. W. (1995). EP component identification and measurement by principal component analysis. Brain and Cognition, 27, 288-310.

Cipollotti, L., Butterworth, B., & Denes, G. (1991). A specific deficit for numbers in case of dense acalculia. Brain, 114(6), 2619-37.

Clearfield, M.W., & Mix, K. (1999). Number versus contour length in infants' discrimination of small visual sets. Psychological Science, 10, 408-411.

Clearfield, M.W., & Mix, K. S. (2001). Amount versus number: Infants' use of area and contour length to discriminate small sets. Journal of Cognition and Development, 2, 243-260.

Cohen Kadosh, R., Lammertyn, J., Izard, V. (2008). Are numbers special? An overview of chronometric, neuroimaging, developmental and comparative studies of mangitude representation. Progress in Neurobiology, 84(2), 132-47.

Cohen-Kadosh, R., Cohen-Kadosh, K., Linden, D. E. J., Gevers, W., Berger, A., & Henik, A. (2007). The brain locus of interaction between number and size: A combined functional magnetic resonance imaging and event-related brain potential study. Journal of Cognitive Neuroscience, 19, 957-970.

Coltheart, M. (1999). Modularity and cognition. Trends in Cognitive Sciences, 3(3), 115-120.

Cooper, R.G.. Jr. (1984). Early number development: Discovering number space with addition and subtraction. In C. Sophian (Ed.), The origins of cognitive skills (pp. 157-192). Hillsdale, NJ: Erlbaum.

Culham, J.C., & Kanwisher, N.G. (2001). Neuroimaging of cognitive functions in human parietal cortex. Current Opinion in Neurobiology, 11(2), 157-63.

De Smedt, B., Janssen, R., Bouwens, K., Verschaffel, L., Boets, B., & Ghesquére, P. (2009b). Working memory and individual differences in mathematics achievement: a longitudinal study. Journal of Experimental Child Psychology, 103(2), 186-201.

De Smedt, B., Verschaffel, L., & Ghesquére, P. (2009a). The predictive value of numerical mangnitude comparison for individual differences in mathematics achievement. Journal of Experimental Child Psychology, 103(4), 469-79.

Dehaene , S. (1996). The organization of brain activations in number comparison: Event-related potentials and the additive factors method. Journal of Cognitive Neuroscience, 8, 47-68.

Dehaene , S. (1997). The number sense: how the mind creates mathematics. New York: Oxford University Press.

Dehaene , S., & Cohen, L. (1995). Towards an anatomical and functional model of number processing. Mathematical Cognition, 1, 83-120.

Dehaene, S., Dehaene-Lambertz, G., & Cohen, L. (1998). Abstract representations of numbers in the animal and human brain. Trends in Neurosciences, 21, 355-361.

Dehaene, S., Piazza, M., Pinel, P., & Cohen, L. (2003). Three parietal circuits for number processing. Cognitive Neuropsychology, 20, 487-506.

Dehaene, S. (1995). The organization of brain activations in number comparison: Event-related potentials and the additive-factors method. Journal of Cognitive Neuroscience, 8:47-68.

Dehaene, S. (1996). The organization of brain activations in number comparison: Event-related potentials and the additive factors method. Journal of Cognitive Neuroscience, 8, 47-68.

Dehaene, S. (1997). The number sense. New York: Oxford University Press.

Dehaene, S., & Akhavein, R. (1995). Attention, automaticity and levels of representation in number processing. Journal of Experimental Psychology: Learning, Memory and Cognition, 21, 314-326.

Dehaene, S., & Cohen, L. (1991). Two mental calculation systems: a case study of severe acalculia with preserved approximation. Neuropsychologia, 29(11), 1045-54.

Dehaene, S., & Cohen, L. (1997). Cerebral pathways for calculation: double dissociation between rote verbal and quantitative knowledge of arithmetic. Cortex, 33, 219-250.

Dehaene, S., & Cohen, L. (1995). Towards an anatomical and functional model of number processing. Mathematical Cognition, 1:83-120, 1995.

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. Journal of Experimental Psychology: General, 122, 371-396.

Dehaene, S., Dehaene-Lambertz, G., & Cohen, L. (1998). Abstract representations of numbers in the animal and human brain. Trends in Neurosciences, 21, 355-361.

Dehaene, S., Molko, N., Cohen, L., & Wilson, A. J. (2004). Arithmetic and the brain. Current Opinion in Neurobiology, 14(2), 218-224.

Dehaene, S., Piazza, M., Pinel, P., & Cohen, L. (2003). Three parietal circuits for number processing. Cognitive Neuropsychology, 20, 487-506.

Dehaene, S., Spelke, E., Pinel, P., Stanescu, R., & Tviskin, S. (1999). Sources of mathematical thinking: Behavioral and brain imaging evidence. Science, 284, 970-974.

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. Journal of Neuroscience Methods, 41, 665-678.

Delorme, A., Sejnowski, T., & Makeig, S. (2007). Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. Neuroimage, 34(4), 1443-9.

Dien, J., & Frishkoff, G. A. (2005). Introduction to principal components analysis of event-related potentials. In T. Handy (Ed.), Event-related potentials: A methods handbook. MIT Press.

Dien, J., Beal, D. J., & Berg, P. (2005). Optimizing principal components analysis of event-related potentials: Matrix type, factor loading weighting, extraction, and rotations. Clinical Neurophysiology, 116(8), 1808-1825.

Dien, J., Spencer, K. M., & Donchin, E. (2004). Parsing the late positive complex: mental chronometry and the ERP components that inhabit the neighborhood of the P300. Psychophysiology, 41(5), 665-678.

Donchin, E. (1966). A multivariate approach to the analysis of average evoked potentials. IEEE Transactions on bio-medical engineering, 13(3), 131-9.

Donchin, E. (1981). Surprise, surprise. Psychophysiology, 18, 493–513.

Donchin, E., & Coles, M. G. H. (1988). Is the P300 component a manifestation of context updating? . Behavioral and Brain Sciences, 11, 357-374.

Donchin, E., Spencer, K. M., & Dien, J. (1978). Cognitive psychophysiology: The endogenous components of the ERP. In E. Callaway, P. Tueting & S. H. Koslow (Eds.), Event-related brain potentioals in man. New York: Academic Press.

Dowker, A. (1997). Young children's addition estimates Mathematical Cognition, 3(2), 141-154.

Droit-Volet, S., Clement, A., & Wearden, J. (2001). Temporal generalization in 3- to 8-year-old children. Journal of Experimental Child Psychology, 80(3), 271-88.

Duffau, H., Denvil, D., Lopes, M., Gasparini, F., Cohen, L., Capelle, L., et al. (2002). Intraoperative mapping of the cortical areas involved in multiplication and subtraction: an electrostimulation study in a patient with a left parietal glioma. J Neurol Neurosurg Psychiatry, 73(6), 733-738.

Duncan, E. M., & McFarland, C. E. (1980). Isolating the effects of symbolic distance and semantic congruity in comparative judgements: An additive factors analysis. Memory and Cognition, 8, 612-622.

Durand, M., Hulme, C, Larkin, R., & Snowling, M. (2005). The cognitive foundations of reading and arithmetic skills in 7- to 10-year-olds. Journal of Experimental Child Psychology, 91, 113-136.

Durston, S., Thomas, K. M., Yang, Y., Ulug, A. M., Zimmerman, R. D., & Casey, B. J. (2002). A neural basis for the development of inhibitory control. Developmental Science, 5(4), 9-16.

Eger, E., Sterzer, P., Russ, M. O., Giraud, A. L., & Kleinschmidt, A. (2003). A supramodal number representation in human intraparietal cortex. Neuron, 37, 719-725.

Eimer, M. (1996). The N2pc component as an indicator of attentional selectivity. Electroencephalography and Clinical Neurophysiology, 99, 225-234.

Fayol, M., Barrouillet, P., & Marinthe, C. (1998). Predicting arithmetical achievement from neuro-psychological performance: a longitudinal study. Cognition, 68(2), B63-B70.

Federmeier, K. D., & Kutas, M. (2002). Picture the difference: electrophysiological investigations of picture processing in the two cerebral hemispheres. Neropsychologia, 40, 730-747.

Feigenson, L., Carey, S., & Hauser, M. (2002b). The representations underlying infants' choice of more: object files versus analog magnitudes. Psychological Science, 13(2), 150-156.

Feigenson, L., Carey, S., & Spelke, E. (2002a). Infants' discrimination of number vs continuous extend. Cognitive Psychology, 44(1), 33-66.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. Trends Cogn Sci, 8(7), 307-314.

Fias, W., Lammertyn, J., Reynvoet, B., Dupont, P., & Orban, G. A. (2003). Parietal representation of symbolic and nonsymbolic magnitude. Journal of cognitive neuroscience, 15, 47-56.

Fischer, M. H., Castel, A. D., Dodd, A. D., & Pratt, J. (2003). Perceiving numbers causes spatial shifts of attention. Nature Neuroscience, 6(6), 555-556.

Fodor, J. (1983) The Modularity of Mind MIT Press Cambridge, Mass.

Fürst, A. J., & Hitch, G. J. (2000). Separate roles for executive and phonological components of working memory in mental arithmetic. Memory & Cognition, 28, 774-782.

Gallistel, C. R. & Gelman, R. (1992). Preverbal and verbal counting and computation. Cognition, 44, 43-74.

Gallistel, C.R., & Gelman, I.I. (2000). Non-verbal numerical cognition: from reals to integeres. Trends in Cognitive Sciences, 4(2), 59-65.

Gazzaniga, M.S., Ivry, R.B., & Mangun, G.B. (2008). Cognitive neuroscience: The biology of the mind. W. W. Norton & Co.

Geary, D. C. (1994). Children's Mathematical Development: Research and Practical Applications. Washington: American Psychological Association.

Geary, D. C. (2004). Mathematics and learning disabilities. Journal of Learning Disabilities, 37, 4-15.

Geary, D. C. (2005). Role of cognitive theory in the study of learning disability in mathematics Journal of Learning Disabilities, 38(4), 305-307.

Geary, D. C., & Hoard, M. K. (2001). Numerical and arithmetical deficits in learning disabled children: Relation to dyscalculia and dyslexia. Aphasiology, 15, 635–647.

Geary, D.C. (1993). Mathematical disabilities: Cognitive, neuropsychological, and genetic components, Psychological Bulletin 114, 2, 345–362.

Geary, D.C. (2007). An evolutionary perspective on learning disability in mathematics. Developmental Neuropsychology, 32(1), 571-519.

Geary, D.C., Hoard, M.K., Byrd-Craven, J., Nugent, L., & Numtee, C. (2007). Cognitive mechanisms underlying achievement deficits in children with mathematical learning disability. Child Development, 78(4), 1343-59.

Gerstadt, C. L., Hong, Y. J., & Diamond, A. (1994). The relationship between cognition and action: Performance of children 3½ -7 years old on a Stroop-like day-night test. Cognition, 53, 129-153.

Gerstmann, J. (1940). Syndrome of finger agnosia, disorientation for right and left, agraphia and acalculia. Archives of Neurology and Psychiatry, 44, 398-408.

Gevers, W., Reynvoet, B., & Fias, W. (2003). The mental representation of ordinal sequences is spatially organized. Cognition, 87(3), B87-95.

Gibbon, J., Church, R.M., & Meck, W.H. (1984). Scalar timing in memory. Annals of the New York Academy of Sciences, 423, 52-77.

Girelli, L., Lucangeli, D., & Butterworth, B. (2000). The development of automaticity in accessing number magnitude. Journal of Experimental Child Psychology, 76, 104-122.

Gobel, S. M., Johansen-Berg, H., Behrens, T., & Rushworth, M. F. (2004). Response-selection-related parietal activation during number comparison. J Cogn Neurosci, 16(9), 1536-1551.

Golub, G.H., & Reinsch, C. (1970). Singular value decomposition and least squares solutions. Numer. Math. 14 (1970) 403–420.

Goodale, M. A., & Millner, A. D. (1992). Separate visual pathways for perception and action. Trends in Neurosciences, 15, 20-25.

Gordon, P. (2004). Numerical cognition without words: Evidence from Amazonia. Science, 306, 496-499.

Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: neural models of stimulus-specific effects. Trends in Cognitive Sciences, 10(1), 14-23.

Grossberg, S., & Repin, D. V. (2003). A neural model of how the brain represents and compares multi-digit numbers: spatial and categorical processes. Neural Netw, 16(8), 1107-1140.

Gross-Tsur, V., Manor, O., & Shalev, R. S. (1996). Developmental dyscalculia: prevalence and demographic features. Developmental Medicine and Child Neurology, 38, 25-33.

Grune, K., Mecklinger, A., & Ullsperger, P. (1993). Mental comparison: P300 component of the ERP reflects the symbolic distance effect. Neuroreport, 4, 1272-1274.

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., Mazzocco, M. M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. Nature Letters, 455, 665-669.

Hamm, J. P., Johnson, B. W., & Kirk, I. J. (2002). Comparison of the N300 and N400 ERPs to picture stimuli in congruent and incongruent contexts. Clinical Neurophysiology, 113, 1339-1350.

Handy, T.C. (2004). Event-related potentials: A methods handbook. MIT Press.

Harner, R.N. (1990). Singulr value decomposition – a general linear model for analysis of multivariate structure in the electroencephalogram. Brain Topography, 3(1), 43-7.

Harner, R.N., & Riggio, S. (1989). Application of singular value decomposition to topographic analysis of flash-evoked potentials. Brain Topography, 2(1-2), 91-8.

Henik, A., & Tzelgov, J. (1982). Is three greater than five: The relation between physical and semantic size in comparison tasks. Memory and Cognition, 10, 389-395.

Hinrichs, J. V., Yurko, D. S., & Hu, J. M. (1981). Two-digit number comparison: Use of place information. Journal of Experimental Psychology: Human Perception and Performance, 7, 890-901.

Hitch, J. G., & McAuley, E. (1991). Working memory in children with specific arithmetical learning difficulties. British Journal of Psychology, 82, 375-386.

Holloway, I. D. & Ansari, D. (2009). Mapping numerical magnitudes onto symbols: The numerical distance effect and individual differences in children's mathematics achievement. Journal of Experimental Child Psychology, 103, 17-29.

Holloway, I.D., & Ansari, D. (2008). Domain-specific and domain-general changes in children's development of number comparison. Developmental Science, 11(5), 664-9.

Hotelling, H. (1933). Simplified calculation of principal components. Psychometrika, 1, 27-35.

Hubbard, E. M., Diester, I., Cantlon, J. F., Ansari, D., Opstal, F., & Troiani, V. (2008). The evolution of numerical cognition: from number neurons to linguistic quantifiers. J Neurosci, 28(46), 11819-11824.

Huizinga, M., Dolan, C. V., & Van Der Molen, M. W. (2006). Age-related change in executive function: Developmental trends and a latent variable analysis. Neuropsychologia, 44, 2017-2036.

Huntley-Fenner, G. & Cannon, E. (2000). Preschoolers' magnitude comparisons are mediated by a preverbal analog mechanism. Psychological Science, 11, 147-152.

Huntley-Fenner, G. (2001). Children's understanding of number is similar to adults' and rats': numerical estimation by 5--7-year-olds. Cognition, 78, B27-B40.

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(51), 19599-19604.

Huttenlocher, J., Jordan, N., & Levine, S. C. (1994). A mental model for early arithmetic. Journal of Experimental Psychology: General, 123, 284-296.

Isaacs, E. B., Edmonds, C. J., Lucas, A., & Gadian, D. G. (2001). Calculation difficulties in children of very low birthweight:A neural correlate. Brain 124, 1701-1707.

Izard, V., Dehaene-Lambertz, G., & Dehaene, S. (2008). Distinct cerebral pathways for object identity and number in human infants. PloS Biology, 6(2), e11.

John, E. R., Ruchkin, D. S., & Villegas, J. (1964). Experimental background: signal analysis and behavioral correlates of evoked potential configuration in cats. Annals of the New York Academy of Sciences, 112, 362-420.

Jordan, K.E., & Brannon, E.M. (2006a). A common representational system governed by Weber's law: nonverbal numerical similarity judgments in 6-year-olds and rhesus macaques. Journal of Experimental Child Psychology, 95(3), 215-29.

Jordan, K.E., & Brannon, E.M. (2006b). Weber's law influences numerical representations in rhesus macaques (Macaca mulatta). Animal Cognition, 9(3), 159-72.

Jordan, K.E., & Brannon, E.M. (2006c). The multisensory representation of number in infancy. Proceedings of the National Academy of Sciences of the U.S.A., 103(9), 3486-9.

Kadosh, R. C., & Henik, A. (2006). Color Congruity Effect: Where do Colors and Numbers Interact in Synesthesia? Cortex, 42(2), 259-263.

Kadosh, R. C., Henik, A., & Rubinsten, O. (2007). The effect of orientation on number word processing. Acta Psychologica, 124(3), 370-381.

Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: object-specific integration of information. . Cognitive Psychology, 24, 175-219.

Karmiloff-Smith, A. (1995). Beyond modularity: Developmental perspective on cognitive science. MIT Press.

Karmiloff-Smith, A. (1997). Beyond modularity: A developmental perspective on cognitive science. Cambridge, MA: MIT Press.

Karmiloff-Smith, A. (2006). The tortuous route from genes to behaviour: A neuroconstructivist approach. Cognitive, Affective and Behavioral Neuroscience, 6(1), 9-17.

Karmiloff-Smith, A. (2009). Nativism versus neuroconstructivism: rethinking the study of developmetal disorders. Developmental Psychology, 45(1), 56-63.

Kaufmann, L., & Nurek, H. C. (2006). Interference effects in a numerical stroopparadigm in 9- to 12- year-old children with ADHD-C. Child Neuropsychology, 12, 223-243.

Kaufmann, L., Koppelstaetter, F., Delazer, M., Siedentopf, C., Rhomberg, P., Golaszewski, S., et al. (2005). Neural correlates of distance and congruity effects in a numerical Stroop task: an event-related fMRI study. Neuroreport, 25, 888-898.

Kaufmann, L., Koppelstaetter, F., Siedentopf, C., Haala, I., Haberlandt, E., Zimerhackl, L. B., et al. (2006). Neural correlates of the number-size interference task in children. Neuroreport, 17, 587-591.

Kawashima, R., Taira, M., Okita, K., Inoue, K., Tajima, N., Yoshida, H., Sasaki, T., Sugiura, M., Watanabe, J., & Fukuda, H. (2004).Cognitive Brain Research, 18, 225-231.

Kayser, J., & Tenke, C.E. (2003). Optimizing PCA methodology for ERP component identification and measurement: theoretical rationale and empirical evaluation. Clinical Neuropsychology, 114(12), 2307-25.

Kayser, J., Tenke, C. E., Gil, R. B., & Bruder, G. E. (2009). Stimulus- and response-locked neuronal generator patterns of auditory and visual word recognition memory in schizophrenia. International Journal of Psychophysiology, 73(3), 186-206.

Koechlin, E., Dehaene, S., & Mehler, J. (1997). Numerical transformations in five-month-old human infants. Mathematical Cognition, 89(3), 89-104.

Koechlin, E., Naccache, L., Block, E., & Dehaene, S. (1999). Primed numbers: Exploring the modularity of numerical representations with masked and unmasked semantic priming. Journal of Experimental Psychology: Human Perception and Performance, 25, 1882-1905.

Krekelberg, B., Boyton, G.M., & van Wezel, R.J. (2006). Adaptation: from single cells to BOLD signals. Trends in Neuroscience, 29(5), 250-256.

Kucian, K., Loenneker, T., Dietrich, T., Dosch, M., Martin, E., & Von Aster, M. (2006). Impaired neural networks for approximate calculation in dyscalculic children: a functional MRI study. Behavioral and Brain Functions, 5, 2-31.

Kutas, M., McCarthy, G., & Donchin, E. (1977). Augmenting mental chronometry: The P300 as a measure of stimulus evaluation in time. Science, 197, 792-795.

Lander, K., Bevan, A., & Butterworth, B. (2004). Developmental dyscalculia and basic numerical capacities: a study of 8-9-year-old students. Cognition, 93(2), 99-125.

Landerl, K., & Kölle, C. (2009). Typical and atypical development of basic numerical skills in elementary school. Journal of Experimental Child Psychology, 103(4), 546-65.

Lehmann, D., & Skrandies, W. (1980). Reference-free identification of components of checkerboard evoked multichannel potential fields. Electroencephalography and Neurophysiology, 48, 609-621.

Levy, L.M., Reis, I.L., & Grafman, J. (1999). Metabolic abnormalities detected b H-MRS in dyscalculia and dysgraphia. Neurology, 53, 639-641.

Libertus, M. E., Pruitt, L. B., Woldorff, M. G., & Brannon, E. M. (2008). Induced alpha-band oscillations reflect ratio-dependent number discrimination in the infant brain. Journal of Cognitive Neuroscience.

Libertus, M. E., Woldorff, M. G., & Brannon, E. M. (2007). Electrophysiological evidence for notation independence in numerical processing. Behavioral and Brain Functions, 3(1).

Lipton, J. S., & Spelke, E. S. (2003). Origins of number sense. Large-number discrimination in human infants. Psychol Sci, 14(5), 396-401.

Logan, G. D. (1988). Toward an Instance Theory of Automatization. Psychological Review, 95(4), 492-527.

Luck, S. J., & Hillyard, S. A. (1994). Electrophysiological correlates of feature analysis during visual search. Psychophysiology, 31, 291-308.

Luck, S.J. (2005). An introduction to the event-related potential technique. MIT Press.

Magliero, A., Bashore, T. R., Coles, M. G., & Donchin, E. (1984). On the dependence of P300 latency on stimulus evaluation processes. Psychophysiology, 21(2), 171-186.

Makeig, S., Westerfield, M., Jung, T.P., Covington, J., Townsend, J., Sejnowski, T.J., & Courchesne, E. (1999b). Functionally independent components of the late positive event-related potential during visual spatial attention. Journal of Neuroscience, 19(7), 2665-80.

Makeig, S., Westerfield, M., Towsend, J., Jung, T.P., Courchesne, E., & Sejnowski, T.J. (1999a). Functionally independent components of early event-related potentials in a visual spatial attention task. Philosophycal Transactions of the Royal Society of London, 354(1387), 1135-44.

Mandler, G., & Shebo, B.J. (1982). Subitizing: an analysis of its component processes. Journal of Experimental Pschology: General, 111(1), 1-22.

Márkus, A. Doktori disszertáció, ELTE, 2005.

Marr, D. (1982). Vision: A computational investigation into the human representation and processing of visual information. San Francisco: W.H. Freeman.

Mazzocco, M. M. M., & Myers, G. F. (2003). Complexities in identifying and defining mathematics learning disability in the primary school-age years. Annals of Dyslexia, 53, 218–253.

McCarthy, G., & Donchin, E. (1981). A metric for thought: a comparison of P300 latency and reaction time. Science, 211, 77-80.

McCloskey, M. (1992). Cognitive mechanisms in numerical processing: evidence from acquired dyscalculia. Cognition, 44(1-2), 107-157.

Mechner, F. (1958). Probability relations within response sequences under ratio reinforcement. Journal of the Experimental Analysis of Behavior, 1, 109-122.

Meck, W. H., & Church, R. M. (1983). A mode control model of counting and timing processses. Journal of Experimental Psychology: Animal Behavior Processes, 9, 320-334.

Menon, V., Rivera, S. M., White, C. D., Glover, G. H., & Reiss, A. L. (2000). Dissociating prefrontal and parietal cortex activation during arithmetic processing. Neuroimage, 4, 357-365.

Mix, K. S. (1999a). Preschoolers' Recognition of Numerical Equivalence: Sequential Sets. Journal of Experimental Child Psychology, 74(4), 309-332.

Mix, K. S. (1999b). Similarity and numerical equivalence: Apperarances count. Cognitive Development, 67, 1592-1608.

Mix, K. S., Huttenlocher, J., & Levine, S. C. (1996). Multiple cues for quantification in infancy: Is number one of them? Psychological Bulletin, 128, 278-294.

Mix, K. S., Huttenlocher, J., & Levine, S. C. (2002). Multiple cues for quantification in infancy: Is number one of them? Psychological Bulletin, 128, 278-294.

Mix, K., Levine, S. C., & Huttenlocher, J. (1997). Numerical Abstraction in Infants: Another Look. Developmental Psychology, 33(3), 423-428.

Molko, N., Cachia, A., Riviére, D., Mangin, J.F., Braundet, M., Le Bihan, D., Cohen, L., & Dehaene, S. (2003). Functional and structural alterations of the intraparietal sulcus in a developmental dyscalculia of genetic origin. Neuron, 40(4), 847-58.

Moyer, R. S., & Landauer, T. (1967). The time required for judgements of numerical inequality. Nature Letters, 215, 1519-1520.

Moyer, R.S. (1973). Comparing objects in memory: evidence suggesting an internal psychophysics. Perception and Psychophysics, 13, 180–184.

Moyer, R.S., & Bayer, R. (1976). Mental comparison of the symbolic distance effect. Cognitive Psychology, 8(2), 228-246.

Mussolin, C., & Noël, M.P. (2007). The nonintentional processing of Arabic numbers in children. Journal of Experimental Neuropsychology, 29(3), 225-34.

Mussolin, C., & Noël, M.P. (2008). Automaticity for numerical magnitude of two-digit Arabic numbers in children. Acta Psychologica, 129(2), 264-72.

Naccache, L., & Dehaene, S. (2001a). The priming method: Imaging unconscious repetition priming reveals an abstract representation of number in the parietal lobes. Cerebral Cortex, 11, 966-974.

Naccache, L., & Dehaene, S. (2001b). Unconscious semantic priming extends to novel unseen stimuli. Cognition, 80, 223-237.

Nieder, A., & Miller, E.K. (2004). A parieto-frontal network for visual numerical information in the monkey. Proceedings of the National Academy of Sciences, 101(19), 7457-62.

Nieder, A., Freedman, D. J., & Miller, E. K. (2002). Representation of the quantity of the visual items in the primate prefrontal cortex. Science, 297, 1708-1711.

Noël, M.P., & Seron, X. (1992). Notational constraints and number processing: A reappraisal of the Gonzalez and Kolers (1982) study. Quarterly Journal of Experimental Psychology, 45, 451-478.

Nunez, P.L., Srinivasan, R. (2006). Electric fields of the brain: The neurophysics of EEG. OUP USA.

O'Hare, A. J., Dien, J., Waterson, L. D., & Savage, C. R. (2008). Activation of the posterior cingulate by semantic priming: A co-registered ERP/fMRI study. Brain Research, 1189, 97-114.

Onton, J., Westerfield, M., Townsend, J., & Makeig, S. (2006). Imaging human EEG dynamics using independent component analysis. Neuroscience and Behavioral reviews, 30(6), 808-22.

Palmeri, T.J., Wong, A.C., & Gauthier, I. (2004). Computational approaches to the development of perceptual expertise. Trends in Cognitive Sciences, 8(8), 378-86.

Passolunghi, M. C., & Siegel, L. S. (2001). Short-term memory, working memory, and inhibitory control in children with difficulties in arithmetic problem solving. . Journal of Experimental Child Psychology, 80, 44-57.

Passolunghi, M. C., & Siegel, L. S. (2004). Working memory and access to numerical information in children with disability in mathematics. Journal of Experimental Child Psychology, 88, 348-367.

Passolunghi, M.C., Cornoldi, C., & De Liberto, S. (1999). Working memory and intrusions of irrelevant information in a group of specific poor problem solvers. Memory and Cognition, 27(5), 779-90.

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. Philosophical Magazine, 2, 559-572.

Piaget, J. (1952). The origins of intelligence in children. International Universities Press.

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.

Piazza, M., Pinel, P., Le Bihan, D., & Dehaene, S. (2007). A magnitude code common to numerosities and number symbols in human intraparietal cortex. Neuron, 53(2), 293-305.

Pica, P., Lemer, C., Izard, V., & Dehaene, D. (2004). Exact and approximate arithmetic in an Amazonian indigene group. Science, 306, 496-499.

Picton, T. W., Bentin, S., Berg, P., Donchin, E., Hillyard, S. A., Johnson, R., Jr., et al. (2000). Guidelines for using human event-related potentials to study cognition: recording standards and publication criteria. Psychophysiology, 37(2), 127-152.

Pinel, P., Dehaene, S., Riviere, D., & LeBihan, D. (2001). Modulation of parietal activation by semantic distance in a number comparison task. Neuroimage, 14, 1013-1026.

Pinel, P., Piazza, M., Le Bihan, D., & Dehaene, S. (2004). Distributed and overlapping cerebral representations of number, size, and luminance during comparative judgments. Neuron, 41, 983-993.

Platt, J. R., & Johnson, D. M. (1971). Localization of position within a homogeneous behavior chain: effects of error contingencies. Learning and Motivation, 2, 386-414.

Posner, M. (1978). Chronometric explorations of mind: Lawrence Erlbaum Associates.

Pourtois, G., Delplanque, S., Michel, C., & Vuilleumier, P. (2008). Beyond conventional event-related brain potential (ERP): Exploring the time-course of visual emotion processing using topographic and principal component analyses Brain Topography, 20, 265-277.

Prevor, M. B., & Diamond, A. (2005). Color-object interference in young children: A Stroop effect in children 3½ - 6½ years old. Cognitive Development, 20, 256-278.

Price, G. R., Holloway, I., Räsänen, P., Vesterinen, M., & Ansari, D. (2007). Impaired parietal magnitude processing in developmental dyscalculia. Current Biology, 17, R1042-1043.

Restle, F. (1970). Speed of adding and comparing numbers. Journal of Experimental Psychology, 83, 274-278.

Reynvoet, B., & Brysbaert, M. (2004). Cross-notation number priming investigated at different stimulus onset asynchronies in parity and naming tasks. Exp Psychol, 51(2), 81-90.

Reynvoet, B., Brysbaert, M., & Fias, W. (2002). Semantic priming in number naming. Quarterly Journal of Experimental Psychology, 55(4), 1127-39.

Reynvoet, B., De Smedt, B., & Van den Bussche, E. (2009). Children’s representation of symbolic magnitude: the development of the priming distance effect. Journal of Experimental Child Psychology, 103(4), 480-9.

Ritter, W., Vaughan, H.G.Jr., & Costa, L.D. (1968). Orienting and habituation to auditory stimuli: A study of short term changes in average evoked responses. Electroencephalography and Clinical Neurophysiology, 10, 550-556.

Rivera, S. M., Menon, V., White, C. D., Glaser, B., & Reiss, A. L. (2002). Functional brain activation during arithmetic processing in females with fragile X syndrome is related to FMR1 protein expression. Human Brain Mapping, 16, 206-218.

Rivera, S. M., Reiss, A. L., Eckert, M. A., & Menon, V. (2005). Developmental changes in mental arithmetic: evidence for increased functional specialization in the left inferior parietal cortex. Cereb Cortex, 15(11), 1779-1790.

Rotzer, S., Kucian, K., Martin, E., Von Aster, M., Klaver, P., & Loenekker, T. (2008). Optimized voxel-based morphometry in children with developmental dyscalculia. Neuroimage.

Rotzer, S., Loenekker, T., Kucian, K., Martin, E., Klaver, P., & von Aster, M. (2009). Dysfunctional neural network of spatial working memory contributes to developmental dyscalculia. Neuropsychologia, 13, 2859-2865.

Rouselle, L., & Noël, M.P. (2007). Basic numerical skills in children with mathematics learning disabilities: a comparison of symbolic vs non-symbolic number magnitude processing. Cognition, 102(3), 361-95.

Rouselle, L., Palmers, E., & Noël, M. P. (2004). Magnitude comparison in preschoolers: What counta? Influence of perceptual variables. Journal of Experimental Child Psychology, 87, 57-84.

Rousselle, L., & Noël, M. P. (2008). The development of automatic numerosity processing in preschoolers: Evidence for Numerosity-Perceptual Interference. Developmental Psychology, 44(2), 544-560.

Rubinsten, O., & Henik, A. (2005). Automatic activation of internal magnitudes: A study of developmental dyscalculia. Neuropsychology, 5, 641-648.

Rubinsten, O., & Henik, A. (2006). Double dissociation of functions in developmental dyslexia and dyscalculia. Educational Psychology, 98, 854-867.

Rubinsten, O., & Henik, A. (2009). Developmental dyscalculia: heterogeneity might not mean different mechanisms. Trends in Cognitive Sciences, 13(2), 92-99.

Rubinsten, O., Henik, A., Berger, A., & Shahar-Shalev, S. (2002). The development of internal representations of magnitude and their association with Arabic numerals. Journal of Experimental Child Psychology, 81(1), 74-92.

Rugg, M.D., & Coles, M.G.H. (Eds) (1996). Electrophysiology of mind: Event-related brain potentials and cognition. Oxford Psychology Series

Rusconi, E., Kwan, B., Giordano, B., Umilta, C., & Butterworth, B. (2005). The mental space of pitch height. Annals of The New York Academy Sciences, 1060; 195-7.

Rusconi, E., Kwan, B., Giordano, B., Umilta, C., & Butterworth, B. (2006). Spatial representation of pitch height: the SMARC effect. Cognition, 99(2), 113-29.

Rusconi, E., Walsh, V., & Butterworth, B. (2005). Dexterity with numbers: rTMS over left angular gyrus disrupts finger gnosis and number processing. Neuropsychologia, 43, 1609-1624.

Schroeter, M. I., Zysset, S., Wahl, M., & Cramon, D. Y. (2004). Prefrontal activation due to Stroop interference increases during development – an event-related fNIRS study. Neuroimage, 23(4), 1317-1325.

Schwarz, W., & Heinze, H. J. (1998). On the interaction of numerical and size information in digit comparison: A behavioural and event-related potential study. Neuropsychologia, 36, 1167-1179.

Schwarz, W., & Ischebeck, A. (2003). On the relative speed account of number - size interferrence in comparative judgements of numerals. Journal of Experimental Psychology: Human Perception and Performance, 29(3), 507-522.

Sekuler, R., & Mierkiewicz, D. (1977). Children's judgment of numerical inequality. Child Development, 48, 630–633.

Shalev, R.S. (2004). Developmental dyscalculia. Journal of Child Neurology, 19(10), 765-71.

Shalev, R.S., & Gross-Tsur, V. (2001). Developmental Dyscalculia. Pediatric Neurology, 24(5), 337-42.

Shalev, R.S., Manor, O., & Gross-Tsur, V. (2005). Developmental dyscalculia: a prospective six-year follow-up. Developmental Medicine and Child Neurology, 47(2), 121-5.

Shalev, R.S., Manor, O., Amir, N., Wertman-Elad, R., & Gross-Tsur, V. (1995). Developmental dyscalculia and brain laterality. Cortex, 31(2), 357-65.

Shalev, R.S., Manor, O., Auerbach, j., & Gross-Tsur, V. (1998). Persistence of developmental dyscalculia: what counts? Results from a 3-year prospective follow-up study. Journal of Pediatrics, 133(3), 358-62.

Shalev, R.S., Manor, O., Kerem, B., Ayali, M., Badichi, N., Friedlander, Y., & Gross-Tsur, V. (2001). Developmental dyscalculia is a familiar learning disability. Journal of Learning Disabilities, 34, 59-65.

Shuman, M., & Kanwisher, N. (2004). Numerical magnitude in the human parietal lobe; tests of representational generality and domain specificity. Neuron, 44(3), 557-69.

Siegel, L.S., & Ryan, E.B. (1989). The development of working memory in normally achieving and subtypes of learning disabled children. Child Development 60(4), 973–980.

Siegler, R. S., & Booth, J. L. (2004). Development of numerical estimation in young children. Child development, 75, 428-444.

Simon , T. J. (1997). Reconceptualizing the origins of number knowledge: A non-numerical account. Cognitive Development, 12, 349-372.

Simon, O., Mangin, J. F., Cohen, L., Bihan, D., & Dehaene, S. (2002). Topographical layout of hand, eye, calculation and language-related areas in the human parietal lobe. Neuron, 33, 475-487.

Simon, T., Hespos, S., & Rochat, P. (1995). Do infants understand simple arithmetic? A replication of Wynn (1992). Cognitive Development, 10, 253-269.

Skrandies, W. (1995). Visual information processing: topography of brain electrical activity. Biological Psychology, 10, 1-15.

Soltész, F. A fejlődési diszkalkulia neuropsichológiai, viselkedéses és elektro-fiziológiai vizsgálata. Szakdolgozat, ELTE, 2004.

Soltész, F., & Szűcs, D. (2009). An electro-physiological temporal principal component analysis of processing stages of number comparison in developmental dyscalculia. Cognitive Development, 24(4), 473-485.

Soltész, F., Szűcs, D., Dékány, J., Márkus, A., & Csépe, V. (2007). A combined event-related potential and neuropsychological investigation of developmental dyscalculia. Neuroscience Letters, 417(2), 181-6.

Spelke, E. S., & Tviskin, S. (2001). Language and number: a bilingual training study. Cognition, 78, 45-88.

Spencer, K. M., Dien, J., & Donchin, E. (1999a). A componential analysis of the ERP elicited by novel events using a dense electrode array. Psychophysiology, 36(3), 409-414.

Spencer, K., Dien, J., & Donchin, E. (2001). Spatiotemporal analysis of the late ERP responses to deviant stimuli. Psychophisiology, 38, 343-358.

Spencer, K.M., Goldstein, A., & Donchin, E. (1999b). What is novel about "novel" stimuli? Effects of event probability on P300 and Novelty P3. Psychophysiology, 36, S11.

Squires, N. K., Squires, K. C., & Hillyard, S. A. (1975). Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. Clinical Neurophysiology, 38, 387-401.

Starkey, P., & Cooper, R. G. (1980). Perception of Numbers by Human Infants. Science, 210, 1033-1034.

Starkey, P., Spelke, E. S., & Gelman, R. (1983). Detection of intermodal numerical correspondences by human infants. Science, 222, 179-181.

Starkey, P., Spelke, E. S., & Gelman, R. (1990). Numerical abstraction by human infants. Cognition, 36, 94-128.

Strauss, M. S., & Curtis, L. E. (1981). Infant perception of numerosity. Child Development, 52, 1146-1152.

Strauss, M. S., & Curtis, L. E. (1984). The development of quantitative knowledge in infancy. In C. Sophian (Ed.), The origins of cognitive skill. Hillsdale, NJ: Erlbaum

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. Journal of Experimental Psychology, 18, 643-662.

Sutton, S., Braren, M., Zubin, J., & John, E. R. (1965). Evoked potential correlates of stimulus uncertainty. Science, 150, 1187-1188.

Swanson, H.L. (1993). Working memory in learning disability subgroups. Journal of Experimental Child Psychology, 56(1), 87-114.

Szűcs , D., & Soltész, F. (2008). The interaction of task-relevant and task-irrelevant stimulus features in the number/size congruency paradigm: An ERP study. Brain Research, 1190, 143-158.

Szűcs, D., & Csepe, V. (2004). Access to numerical information is dependent on the modality of stimulus presentation in mental addition: a combined ERP and behavioral study. Brain Res Cogn Brain Res, 19(1), 10-27.

Szűcs, D., & Csepe, V. (2005). The effect of numerical distance and stimulus probability on ERP components elicited by numerical incongruencies in mental addition. Brain Res Cogn Brain Res, 22(2), 289-300.

Szűcs, D., & Soltész, F. (2007). Event-related potentials dissociate facilitation and interference effects in the numerical Stroop paradigm. Neuropsychologia, 45(14), 3190-3202.

Szűcs, D., Soltész, F., Jármi, É., & Csépe, V. (2007). The speed of magnitude processing and executive functions in controlled and automatic number comparison in children: an electro-encephalography study. Behavioral and Brain Functions, 3 (23), 1-20.

Tabachnik, B. G., & Fidell, L. S. (2007). Using multivariate statistics. USA: Pearson Education.

Temple, E., & Posner, M. I. (1998). Brain mechanisms of quantity are similar in 5-year-old children and adults. Proceedings of the National Academy of Sciences of the USA (PNAS), 95, 7836-7841.

Trick, L.M., & Pylysyn, Z.W. (1994). Why are small and large numbers enumerated differently? A limited-capacity preattentive stage in vision. Psychological Review, 101(1), 80-102.

Tudusciuc, O., & Nieder, A. (2007). Neuronal population coding of continuous and discrete quantity in the primate posterior parietal cortex. Proceedings of the National Academy of Sciences, 104(36), 14513-8.

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.

Uller, C., Carey, S., Huntley-Fenner, G. & Klatt, L. (1999). What representations might underlie infant numerical knowledge. Cognitive Development, 14:1, 1-36.

van Loosbroek, E., & Smitsman, W. (1990). Visual perception of numerosity in infancy. Developmental Psychology, 26, 916-922.

vanMarle, K., & Wynn, K. (2006). Six-month-old infants use analog magnitudes to represent duration. Dev Sci, 9(5), F41-49.

Venkatraman, V., Siong, S.C., Chee, M.W., & Ansari, D. (2006). Effect of language switching on arithmetic: a bilingual fMRI study. Journal of Cognitive Neuroscience, 18(1), 64-74.

Vogel, E. K., & Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. Nature, 15, 748-751.

Vogt, B. A., Vogt, L., & Laureys, S. (2006). Cytology and functionally correlated circuits of human posterior cingulated areas. Neuroimage, 29, 425-466.

Von Aster, M.,G., & Shalev, R.S. (2007). Number development and developmental dyscalculia. Developomental Medicine and Child Neurology, 49(11), 868-73.

Walsh, V. (2003). A theory of magnitude: Common cortical metrics of time space and quantity. Trends in Cognitive Sciences, 7(11), 483-488.

Whalen, J., Gallistel, C. R., & Gelman, R. (1999). Nonverbal counting in humans: The psychophysics of number representation. Psychological Science, 10, 130-137.

Wood, J. N., & Spelke, E. S. (2005). Infants' enumeration of actions: numerical discrimination and its signature limits. Dev Sci, 8(2), 173-181.

Wynn, K. (1990). Children’s understanding of counting. Cognition, 36, 155-193.

Wynn, K. (1992). Addition and subtraction by human infants. Nature, 27(358), 749-50.

Wynn, K., Bloom, P., & Chiang, W. C. (2002). Enumeration of collective entities by 5-month-old infants. Cognition, 83, B55-62.

Xu, F. (2003). Numerosity discrimination in infants: Evidence for two systems of representations. Cognition, 89, B15-B25.

Xu, F., & Arriaga, R.I. (2007). Number discrimination in 10-month-old infants. British Journal of Developmental Psychology, 25, 103-8.

Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. Cognition, 74(1), B1-B11.

Xu, F., Spelke, E., & Goddard, S. (2005). Number sense in human infants. Developmental Science, 8, 88-101.

Zbrodoff, N. J., & Logan, G. D. (1986). On the autonomy of mental processes: A case study of arthimetic. Journal of Experimental Psychology: General, 115(2), 118-130.

Zorzi, M., Prifits, K., & Umilta, C. (2002). Neglect disrupts the mental number line. Nature, 417, 138.