The influence of the place-value structure

of the Arabic number system on

two-digit number processing

–

Representational, developmental,

neuropsychological and computational aspects

Dissertation
der Fakultät für Informations- und Kognitionswissenschaften
der Eberhard-Karls-Universität Tübingen
zur Erlangung des Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)

Vorgelegt von

Dipl.-Psych., M. Sc. Korbinian Möller

aus München

Tübingen

2010

Tag der mündlichen Qualifikation:	14. Juli 2010
Dekan:	Prof. Dr. Oliver Kohlbacher
1. Berichterstatter:	Prof. Dr. Hans-Christoph Nürk
2. Berichterstatter:	Prof. Dr. Karin Landerl

Ich möchte mich an dieser Stelle bei meinem Doktorvater
Prof. Dr. Hans-Christoph Nürk
für die hervorragende fachliche Betreuung meiner Arbeit bedanken. Aus der gemeinsamen Arbeit, durch Diskussionen zum Thema sowie seinen zahlreichen Anmerkungen und Kommentaren zu ersten Teilen dieser Arbeit durfte ich viel lernen, auch, wenn sie mit dazu beigetragen haben, dass die vorliegende Arbeit ihren jetzigen Umfang erhielt.

Gedankt sei auch meinen Kollegen, die mich nicht nur unterstützten durchzuhalten, sondern auch an der einen oder anderen Studie beteiligt gewesen sind. Sie haben diese Arbeit erst mit ermöglicht. Mein Dank gilt euch Silvia, Elise, Anne, Ursula, Julia und Stefan.

Zuletzt gebührt einer Frau Dank, die selbst kein Wort und keine Zeile zu dieser Arbeit beigetragen, ohne deren Worte, deren Verständnis, deren Herz, kurz, ohne deren Liebe es diese Dissertation in dieser Form jedoch nicht gegeben hätte.

Fabienne,

Wenn Deine Blicke Bände sprechen,
– ein Lächlen Deine Lippen ziert
und Deine Worte daran brechen,
mein Traum im Glanze Deiner Augen sich verliert.

Machst mich mit Deinem Antlitz schweigen,
und doch begeisterst mich so sehr,
wenn Tag und Abend sich schon neigen,
fragt sich, was will ich mehr?

PRÉCIS

In recent years, research on the human representation of number, its development, and its related capabilities (e.g., mental arithmetic) has flourished. A latest model of numerical processing, the Triple Code Model (Dehane & Cohen, 1995; 1997) is even quite explicit on the involved representations (i.e., visual number form, rote verbal knowledge, and an analogue magnitude representation) and their associated neural correlates. However, the dominating organizational principle of our Arabic number system, i.e., its place-value structure, has received much less interest. This is reflected by the lack of a systematic evaluation of the influence of the place-value structure of the Arabic number system on number processing. The current thesis aimed at making a first step towards a more elaborate understanding of place-value influences as regards its representation in adult number processing, its influence on the development of numerical representations as well as arithmetic capabilities, and its neuro-cognitive underpinnings. Consolidated evaluation of the results of the individual studies provided converging evidence that place-value constraints exhibit a task invariant influence on human numerical cognition and its development suggesting that (i) the mental number line is not analogue holistic in nature but incorporates place-value information as well, (ii) place-value processing seems to develop culture invariant but is nevertheless influenced by a language's number word system, and (iii) on a neurofunctional level specific neural correlates of processing place-value information can be identified. Evidence for these conclusions comes from a variety of tasks administred to different populations and assessed using differing methodologies (including eye-tracking and fMRI). Taken together, the results obtained by the current thesis can be recapitulated as indicating a comprehensive, culture invariant but linguistically influenced and developmentally relevant influence of the place-value structure of the Arabic number system on human numerical cognition.

ZUSAMMENFASSUNG

Das wissenschaftliche Interesse an der menschlichen Fähigkeit zur Repräsentation und Verarbeitung von Zahlen ist in den letzten Jahren stärker geworden. Die beteiligten Repräsentationen wie z.B. die der visuellen Zahlenform, des verbalen Faktenwissens und der analogen Größenrepräsentation, sowie deren neuronalen Korrelate werden in einem aktuellen Modell der Zahlenverarbeitung (z.B. Triple Code Model, Dehaene & Cohen, 1995; 1997) bereits sehr genau beschrieben. Jedoch fehlt bisher eine systematische Untersuchung zum Einfluss des dominierenden Organisationsprinzips unseres arabischen Zahlsystems, dem Platz x Wert System, auf die Zahlenverarbeitungfähigkeit. Die vorliegende Studie stellt dabei einen ersten Schritt in Richtung eines besseren Verständnisses von Platz x Wert Einflüssen, deren Auswirkungen auf die Endwicklung der Zahlenverarbeitungsfähigkeiten bei Kindern und ihrer neurokognitiven Grundlagen dar. Zusammenfassend lässt sich sagen, dass die einzelnen Studien der vorliegenden Arbeit übereinstimmend zeigen, dass Aufgaben übergreifende Einflüsse des Platz x Wert Systems bei Erwachsenen und Kindern nahelegen, dass (i) der mentale Zahlenstrahl Informationen zur Platz x Wert Kodierung von mehrstelligen Zahlen beinhaltet, (ii) dass sich die Verarbeitung von Platz x Wert Information kulturinvariant zu entwickeln scheint, aber doch Spracheinflüssen unterliegt, und (iii) dass sich auf neurofunktioneller Ebene spezifische neuronale Korrelate der Verarbeitung von Platz x Wert Information identifizieren lassen. Ergebnisse, die diese Schlussfolgerungen untermauern kommen dabei von einer Vielzahl numerischer Aufgaben, die von unterschiedlichen Stichproben bearbeitet wurden und deren Ergebnisse mittels verschiedener experimenteller Methoden erhoben wurden (u.a., Eye-Tracking und fMRT). Grundsätzlich lassen diese Ergebnisse den Schluss auf unfassende, Kultur übergreifende, aber von Sprache moderierte und entwicklungsrelevante Einflüsse der Platz x Wert Struktur des arabischen Zahlsystems auf die menschliche Fähigkeit zur Zahlenverarbeitung zu.

TABLE OF CONTENTS

Section 1: General Introduction

<u>Section 2: On task invariant influences of the place-value structure of the Arabic umber system on numerical cognition</u>

Study 1: Internal number magnitude representation is not holistic, either

Study 2: On the cognitive instantiation of the carry effect in mental addition – Evidence from eyetracking

Section 3: On the importance of place-value integration in arithmetical development

Study 3: Language effects on children's non-verbal number line estimations

Section 4: The neurocognitive underpinnings of place-value integration

Study 5: All for one but not one for all: How multiple number representations are recruited in one numerical task

Study 6: Impairments of the mental number line for two-digit numbers in neglect

Section 6: Summary

Section 7: General Discussion

LIST OF FIGURES

Section 1: General Introduction

Section 2: On task invariant influences of the place-value structure of the Arabic umber system on numerical cognition

Study 1: Internal number magnitude representation is not holistic, either

Study 2: On the cognitive instantiation of the carry effect in addition – Evidence from eyetracking

Section 3: On the importance of place-value integration in arithmetical development

Study 3: Language effects on children's non-verbal number line estimations

Study 4: Early place-value understanding as a precursor for later arithmetic performance – a longitudinal study on numerical development

Section 4: The neurocognitive underpinnings of place-value integration

Study 5: All for one but not one for all: How multiple number representations are recruited in one numerical task

Study 6: Impairments of the mental number line for two-digit numbers in neglect

Section 5: A computational model of place-value integration in two-digit number comparison

Study 7: Two-digit number processing – holistic, decomposed or hybrid? A computational modelling approach

Section 7: General Discussion

LIST OF TABLES

Section 1: General Introduction

Section 3: On the importance of place-value integration in arithmetical development

Study 4: Early place-value understanding as a precursor for later arithmetic performance – a longitudinal study on numerical development

Section 4: The neurocognitive underpinnings of place-value integration

Study 5: All for one but not one for all: How multiple number representations are recruited in one numerical task

Study 6: Impairments of the mental number line for two-digit numbers in neglect

Section 5: A computational model of place-value integration in two-digit number comparison

Study 7: Two-digit number processing – holistic, decomposed or hybrid? A computational modelling approach

Section 1

General Introduction

Number, the simplest and most universal idea;

for number applies itself to men, angels, actions, thoughts,

– every thing that either doth exist or can be imagined.

(John Locke, 1689)

GENERAL INTRODUCTION

No other set of symbols is so widely spread in its use and understanding as the Hindu-Arabic or Western number system (henceforth referred to as Arabic number system). Its constituting digits (i.e., 0, 1, 2, 3, 4, 5, 6, 7, 8, and 9) are known to and understood in their meaning by literate but also a lot of illiterate individuals all over the world, thereby, being more prevalent and more universally employed than any language or script.

Nevertheless, the Arabic number system is only one out of the approximately 100 systems of numerical notation known. According to Chrisomalis (2004, p. 38) "a numerical notation system is a visual but primarily non-phonetic structured system for representing numbers". Thus, numerical notation systems can be clearly differentiated from lexical numeral systems reflecting the number words used to transcode symbolic digital numerical input into spoken or written number words (henceforth referred to as number word system). Obviously, there are specific correspondences between the numerical notation system and the lexical numerical system. For instance, in the herodianic Greek number system the symbols representing 5, 10, 100, 1000, and 10,000 were derived from the first letter from their corresponding number words (see Löffler, 1912 for a more detailed dicussion). However, unlike digital numerical notation systems such as the Arabic number system, number word systems differ between languages and even within the same language different regional influences can be observed. In this context it is interesting that the left-to-right order of tens and units in two-digit numbers is not retained in their corresponding number words in all languages (e.g., 27 → twenty seven), rather the units are spoken before the tens in some languages' number words (e.g., 27 → achtundzwanzig in German [seven and twenty]). This so-called inversion property is found among others in German, Maltese, Arabic, Norwegian, etc. (see Comrie, 2005 for an overview). On the other hand, there are also intra-language variations. For instance, in the Czech language the non-inverted number words are officially used and taught at school. Nevertheless, in some parts of Czech inverted number words are primarily used in everyday life (see Pixner, 2009; see also Seron, Deloche, & Noël, 1992; Seron & Fayol, 1994 for differences in French number words used in Belgium and France). Taken together, these examples illustrate quite nicely that although there are correspondences between the numerical notation system and the number words used to vocalize symbolic numbers, most number word systems involve a lot irregularities that are not inherent in the numerical notation system they actually refer to (i.e., the Arabic number system).

In this vein, the generality of the Arabic number system is striking. In the following, a typology of numerical notation systems shall be introduced (cf. Chrisomalis, 2004). Subsequently, the most important characteristics of the currently dominating Arabic number system shall be outlined.

The Arabic number system in a typology of notational systems

Apart from very simple tally-like notational approaches relying on one-to-one correspondence between the to-be-counted objects and an equal number of markings or body parts all numerical notation systems have to be structured in a specific way as this is a prerequisite of their actual function: to represent any given number (in the best case) by a limited set of symbols. To achieve this goal numerical notation systems are dimensionally organized into bases and powers, for instance, the Arabic number system is a base-10 system. In all notational systems the powers of the base are of special designation (as can be mathematically indicated by an exponent $10^0 = 1$; $10^1 = 10$, $10^2 = 100$ etc. for the Arabic number system). However, please note that there are also numerical notation systems with more than one base for which a comprehensive mathematical description is more complex (e.g., the Roman notation system with the main base 10 and its power levels, i.e., 10^0 = I; 10^1 = X, 10^2 = C etc. and the sub-base 5 and its power levels $5^1 \times 10^0$ = V; $5^1 \times 10^1$ = L, $5^1 \times 10^2$ = D). When expressing e.g., the number of items in a set by their corresponding number there are different possibilities to code the quantity to-be-represented at each individual power level but also in the way the single power levels are integrated into a coherent overall representation of a given number are possible. In his recent typology for numerical notation, Chrisomalis (2004) used exactly such intra- as well as interexponential (referring to the exponent coding the power dimension) characteristics of notation systems to develop a clearly structured typology (see also Guitel, 1975 as well as Zhang & Norman, 1995 for possible typologies relying on different structural aspects of numerical notation systems). Both of these dimensions differentiate between a number of principles determining the way numerical representation is organized.

According to Chrisomalis (2004) the intraexponetial dimension can be subdivided into cumulative, ciphered, and multiplicative organization subtypes. In systems with a cumulative intraexponential structure any power of the base is indicated by repeating number symbols until the sum of these symbols equals the to-be-represented quantity at the particular power level (e.g., CCC represents 300 in the Roman number system, literally meaning 100 + 100 + 100). Contrarily, in ciphered numerical notation systems each power level is represented by only one number symbol. Thus, coding a given number requires as many number symbols as the number involves power levels. Thus, to transcode sixty eight into the symbolic notation two number symbols are necessary, one for the 10^0 and another one for the 10^1 power level. Finally, multiplicative intraexponential organization is indicated when the representation of each power level incorporates two number symbols: a power sign reflecting the actual power level, and a unit sign specifying the quantity at the respective power level (e.g., the traditional Chinese number system, see Table 1).

On the second dimension interexponential organization can be separated into additive and positional structuring principles. In the latter, positionally structured, notational systems the value of a particular number symbol is principally defined by its position within the string of symbols representing a given number. For example, in the positional Arabic number system the digit 3 in 36 reflects a quantity different from that represented by the same digit in 63 because they hold different positions, thereby, reflecting different power levels. On the other hand, in notational systems with an additive interexponential structure the total value of the number can be achieved by just adding up the values quantified at each individual power level. The Roman notational system is the best known example of additive interexponential structuring (see Table 1).

As can easily be derived from the two-dimensional typology of notational systems introduced above the Arabic number system is a ciphered positional notation system – ciphered as each power level is represented by just one number symbol and positional as it has a clear place-value structure with the value of the digits being unambiguously determined by their position within the digit string. Thereby, place-value coding is much more parsimonious than e.g., the Roman notation system in which different values (in particular the powers of the bases) are coded by different symbols instead of using the same set of symbols.

From Table 1 it can be observed that there are or have been numerical notation systems complying with the constraints of almost every possible combination of intra- and interexponential structuring principles. The only exception is presented by the combination of positional and multiplicative structuring which is just logically impossible. A number system in which the quantity at a certain power level and this power level itself are represented by two distinct symbols (i.e., multiplicative organization) cannot comply with the limitation of coding the value of a number symbol not only by its shape but also by its position within the string of number symbols (i.e., positional structuring). In a positional system the unit sign and the power sign used for multiplicative coding of one power level would rather reflect two adjacent power levels, thereby making the combination of multiplicative and positional coding impossible.

Table 1: Intra- and interexponential structuring and their combinations in known numerical notation systems

		Interexponetial Structure	
		Additive Power levels are summed up to obtain total value	**Positional** Value of power levels multiplied with positionally coded multiplier (e.g., hundreds) before summation
Intraexponential structure	**Cumulative** Sum of signs per power value indicate total value	**Roman numerals** 678 = DCLXXVIII (500+100+50+10+10+5+1+1+1)	**Babylonian cuneiform** 678 =  (10 + 1) x 60 + (10 +8) x 1
	Ciphered Total value represented by only one sign	**Greek alphabetic** 678 $\bar{\chi}\bar{o}\bar{\eta}$ = (600 + 70 + 8)	**Arabic number system** 678 (6 x 100 + 7 x 10 + 8)
	Multiplicative Total value computed by multiplication of unit sign(s) and a power sign	**Traditional Chinese** 678 六百七十八 = 6 100 7 10 8	**Logically excluded**

From the fact that the Arabic number system is by far the dominating numerical notation system one may conclude that "the evolution of written numeration converges [in the Arabic number system] […] because place-value coding is the best available notation" as

Dehaene (1997, p. 101) claims (see also Boyer, 1944 for discussion of this point). On the one hand, such an assumption is corroborated by considering the historical development of numerical notation systems. In a detailed evaluation of transformation processes of numerical notation systems Chrisomalis (2004) found that cumulative notation systems did not develop from by non-cumulative antecessors and positional notation systems were never followed up by additive ones. However, on the other hand, Chrisomalis (2004) also noted that knowing which systems were used and then replaced by which other system is only part of the story. It also has to be considered for which purposes (i.e., administrative or technical needs, formal mathematics) a notational system was used and how good the different structural and organizational aspects of the respective notation system coped with the related requirements. In this vein, the Arabic number system with its characteristic place-value structure may be one of the (maybe the) best-suited notation systems for the current state in science, economics, and everyday life. Convergent evidence for this argument comes from Zhang and Norman (1995) who evaluated the applicability of different numerical notation systems by the degree to which they allow for an external representation of information required to perform arithmetic procedures such as multi-digit multiplication. In this context, external means that the respective information can be read from the symbolic notation of a given number while information being exclusively represented internally has to be retrieved from memory. For the case of the Arabic number system this means that e.g., the power levels are represented externally as indicated by the position a certain digit holds within a given number. On the other hand, the semantic meaning of the individual digits, i.e., the quantity they represent, has to be retrieved from memory as the shape of the digits does not directly code the quantity they represent as e.g., the number of marks in any tally-like system does. When analyzing the relation of externally to internally available information needed to multiply two multi-digit numbers Zhang and Norman (1995) proposed that such multiplication involves six steps [i.e.,

(i) separating power and base dimensions, (ii) get base values, (iii) multiply base values, (iv) get power values, (v) add power values, and (vi) attach power values to the product of the base values] of which four [i.e., steps (i), (iv), (v), and (vi)] and thus more than in any other numerical notation system are based on externally available information in the Arabic number system. Based on this observation Zhang and Norman (1995) argue that such comprehensive use of external representations by the Arabic number system may be the most important reason why it is so widely established and has been recognized as on of the most genius inventions in human history. Nevertheless, it is almost impossible to say whether future developments may or may not come up with requirements incompatible with the structuring of the Arabic number system, thereby, leading to transformational processes or even to its replacement.

Yet, for the moment the Arabic number system is a vital part of almost everybody's everyday life. Being numerate (in addition to being literate) is already an integral prerequisite of managing one's life at the beginning of the 21st century. Indeed, Bynner and Parsons (1997, see also Parsons & Bynner, 2005) found insufficient numeracy to be a more pronounced impairment regarding job and promotion prospects as compared to poor literacy. Against this background an understanding of how numerical information is cognitively represented complying with the specifications of the Arabic number system in general and its structural place-value organization in particular is of specific importance to corroborate elaborate and efficient teaching and training of numerical/mathematical competencies (see above for the representational advantages of the Arabic number system). As a first step it may be interesting to evaluate in what way the representation of the place-value structure of the Arabic number system is implemented in current models of numerical cognition. The following section will address this issue.

Arabic place-value coding in current models of number processing

In the following, the two models of numerical cognition by McCloskey (1992) and Dehaene and Cohen (1995; 1997) shall be described and evaluated by their capability to represent the place-value structure of the Arabic number system.

The Abstract Code Model

The Abstract Code model (McCloskey, 1992; McCloskey & Macaruso, 1994, 1995, see Figure 1) proposes that its subsystems (i.e., comprehension, calculation and response production) communicate through a single, abstract, semantic quantity code. The comprehension subsystem transforms different numerical inputs into the abstract code on which calculation and response generation subsequently operate. This abstract code is seen as a decomposition of numbers into their powers of ten complying with the place-value structure of the Arabic number system (e.g. 358 is represented as: $\{3\}\ 10^2\ \{5\}\ 10^1\ \{8\}\ 10^0$). Access to this abstract code is necessarily required before any other numerical process is possible. The calculation subsystem operates only on this code and is supposed to include a memory of basic number facts and rules, such as $2 * 4 = 8$, and $N * 0 = 0$, respectively. Finally, the production subsystem transcodes the abstract code back into Arabic, written or spoken verbal number format as required by the current task. To account for observed dissociations of number naming (i.e., transcoding) and calculations in neuropsychological patients Cipolotti and Butterworth (1995) proposed an additional asemantic route involved in transcoding numbers from one notation into another which bypasses the central abstract semantic quantity code (see also Ablinger, Weniger, & Willmes, 2006 for an intervention study on transcoding performance in an aphasic patient with spared calculation ability).

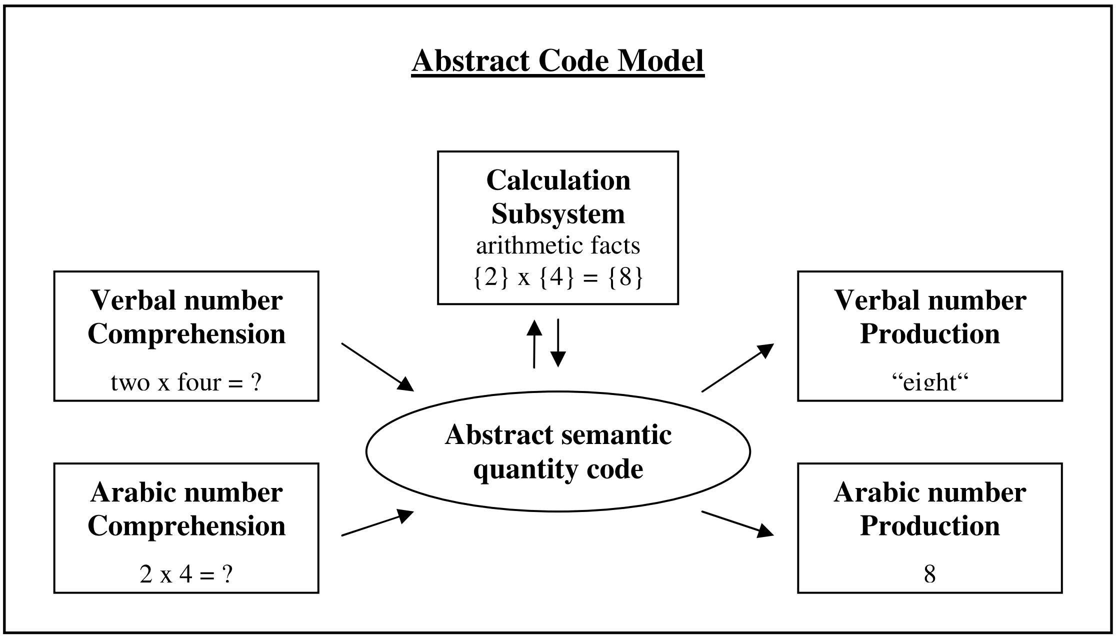


Figure 1: *McCloskey's Abstract Code Model (adapted from Cohen & Dehaene, 1995, p.132)*

The Triple Code Model

Dehaene's (1992, Dehaene & Cohen, 1995, 1997; see also Dehaene, Piazza, Pinel & Cohen, 2003 and Hubbard, Piazza, Pinel, & Dehaene, 2005 for model specifications) Triple Code Model postulates three codes upon which number processing is based: (i) an analogue holistic magnitude representation, (ii) a visual-Arabic number form, and (iii) an auditory-verbal code (see Figure 2). In contrast to the Abstract Code Model (McCloskey, 1992) these three codes can directly activate one another without the central bottleneck of an amodal, semantic code. Thus, if a problem is encountered in a non-specialised modality the information will be transcoded directly to the relevant modality for further processing (see also Cipolotti & Butterworth, 1995 for a similar proposal extending McCloskey's abstract code model). On the other hand, and in accordance with McCloskey's model, task specificity is assumed for each of the components of the Triple Code Model: For instance, any kind of quantity comparison and approximate calculation are believed to mainly rely on the analogue magnitude code. From the beginning the Triple Code Model was conceptualized as a neuro-functional model, thereby, intended to make predictions about where in the human brain the three representations proposed are to be localized. While the neuro-anatomical correlate of the analogue magnitude representation remained quite vague in the first compositions of the model (Dehaene & Cohen, 1995; 1997), Dehaene and colleagues (2003) circumscribed the cortex sites assumed to be specifically involved in magnitude processing more precisely: On the one hand, the core processing of quantity information is proposed to be subserved by the horizontal parts of the intraparietal sulcus (IPS) as an abstract and notation invariant representation of magnitude. This magnitude representation is suggested to be spatially organized by means of a left-to-right oriented mental number line. When processing magnitude information the corresponding parts of this mental number line are assumed to be activated. Thereby, performing numerical tasks as basic as magnitude comparisons requires attentional navigating along the mental number line to integrate the necessary information. Dehaene and colleagues (2003) postulated such processes of mental navigation calling upon the spatial attributes of the mental representation of number magnitude to be subserved by posterior superior parts of the parietal lobe (PSPL) adjacent to the (IPS). Thus, in its latest version the Triple Code Model specifically includes a spatial representation of number magnitude apart from an abstract magnitude code. However, please note that generally magnitude representation in the Triple Code Model is assumed to be holistic in nature, meaning that multi-digit numbers are represented as one integrated entity rather than decomposed into their constituting digits. Thus, the place-value structure of the Arabic number system is not retained in the magnitude representation postulated by the Triple Code Model. Finally, the neuro-anatomical correlates of the auditory-verbal representation (including arithmetic, e.g., multiplication fact knowledge) as well as the visual Arabic number form are suggested to be found in the left angular gyrus and the temporo-occipetal junction respectively. Yet, as the current thesis is primarily concerned with aspects of the magnitude representation and in particular with evaluating the representation of the place-value structure of the Arabic number system within the representation of number magnitude, the neuro-functional implications of the auditory-verbal and the visual number form representation will not be specified any further.

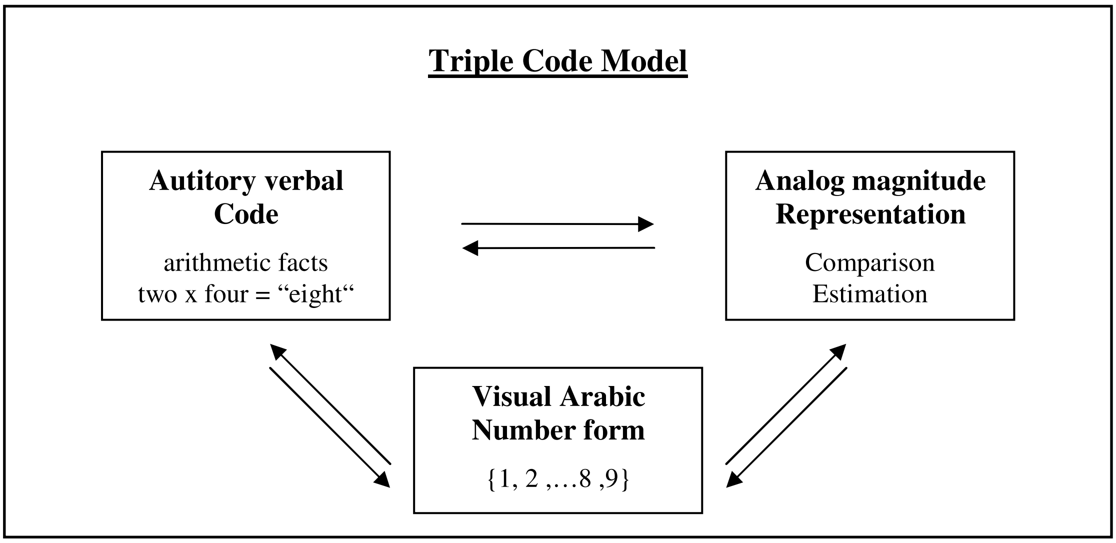


Figure 2: *Dehaene & Cohen’s (1995) Triple Code Model (adapted from Cohen & Dehaene, 1995, p. 132)*

In summary, it is evident, that the two models differ substantially in respect of the implementation of the place-value structure of the Arabic number system. On the one hand, an amodal semantic representation of number magnitude is at the heart of McCloskey’s model and this magnitude representation is organized along place-value constraints. On the other hand, within the framework of the Triple Code Model place-value structuring is not assumed for the representation of analogue magnitude. Instead, the magnitude of any number is represented by a holistic code not including place-value information. Anyway, there is some inconsistency within the Triple Code Model regarding the influence of place-value information. For the case of multi-digit arithmetic, in particular addition and subtraction Dehaene and Cohen (1995) propose a solution process based on repetitively solving single digit problems at corresponding place-value positions, i.e., at the same power level [e.g., $24 + 61 = (2 + 6) * 10^1 + (4 + 1) * 10^0$].

However, at this point it has to be emphasized that although sufficiently implementing the place-value structure of the Arabic number system, the model by McCloskey and Colleagues (McCloskey, 1992; McCloskey & Macaruso, 1994; 1995) is more or less disproved. In particular, the assumption of a central representation of number magnitude upon which all numerical operations such as transcoding, calculation, comparison, etc. operate was not supported by empirical evidence. In this respect the Triple Code Model by Dehaene and Cohen (1995; 1997) is the more powerful and influential model that is largely backed by behavioural data, neuropsychological patient data, and neuro-imaging data (see Dehaene et al., 2003; Dehaene, 2009 for a reviews). For the current thesis two specificities of the Triple Code Model will be of particular relevance. On the one hand, the Triple Code Model serves as a starting point for any functional interpretation of the observed brain activation pattern for any neuro-cognitve / neuro-imaging evaluation of the underlying processes in numerical cognition such as the number bisection task (see Study 5 of the current thesis) as it allows for functional and neuro-anatomical predictions. On the other hand, and even more importantly, the assumption of a holistic analogue representation of number magnitude was the dominant view in research of numerical cognition and is still promoted (e.g., Dehaene, Dupoux, & Mehler, 1990; Ganor-Stern, Pinhas, & Tzelgov, 2009; Hinrichs, Yurko, & Ho, 1981; Moyer & Landauer, 1967; Poltrock & Schwartz, 1984; Restle, 1970; Zhou, Chen, Chen, & Dong, 2008).

An alternative account: decomposed processing of tens and units

The assumption of holistic number magnitude representation (Dehaene & Cohen, 1995; 1997) was also questioned in recent years. In 2001 Nuerk and colleagues (Nuerk, Weger, & Willmes, 2001) observed the so-called unit-decade compatibility effect indicating decomposed processing of tens and units in two-digit number comparison. Unlike most previous studies on two-digit number comparison Nuerk et al. (2001) did not use a fixed standard (e.g., 55 or 66, cf. Dehaene et al., 1990) to which a presented number had to be compared. Instead, participants had to single out the larger number of a pair of two-digit numbers. Generally, when comparing two two-digit numbers separate comparisons of tens and units may result in either compatible or incompatible decision biases. For instance, in the number pair 38_53 the larger number contains the smaller unit digit. Thereby, although the overall relation also holds for the comparison of the tens digits it this is not the case for comparing the unit digits (i.e., 28 < 53 and 2 < 5, but 8 > 3). Thus, separate comparisons of tens and unit result in antidromic, i.e., incompatible decision biases. On the other hand, in a number pair such as 42_57 the decision biases for the separate comparisons of tens and units are both congruent with the overall decision (i.e., 42 > 57, 4 < 5, and 2 < 7). Nuerk and co-workers observed that although overall distance was held constant between compatible and incompatible number pairs (e.g., 15 in both examples above) incompatible number pairs were followed by longer response latencies and more errors than their compatible counterparts. As overall distance was matched between compatible and incompatible number pairs, no compatibility effect should be obtained if an exclusively analogue (holistic) magnitude representation were engaged. Replications of the compatibility effect for Arabic stimuli arranged in different layouts (Nuerk, Weger, Willmes, 2004a; Ratinckx, Nuerk, van Dijk & Willmes, 2006) and also for number words (Nuerk, Weger & Willmes, 2002a; 2005; Macizo, & Herrera, in press) indicated that it is not a purely perceptual effect. Additionally, interactions of the compatibility effect with distance measures of tens and units implicated that it is not a common attentional congruity effect (i.e., large unit distances were associated with a more pronounced compatibility effect for RTs, whereas the same relation was observed for small decade distances and the compatibility effect for error rates). In other words, the results cannot be explained by exclusively assuming a response conflict because unit and decade distance must be processed separately to result in the antidromic influence just mentioned (see also Nuerk & Willmes, 2005 for a more detailed discussion of this point.

Yet, to account for the compatibility effect a model is necessary that allows us to differentiate between processing the tens digits and processing the unit digits complying with the place-value structure of the Arabic number system: a strictly decomposed or at least a hybrid model of two-digit number representation. In the strictly decomposed model the magnitudes of tens and units are represented separately (possibly on concomitant number lines, cf. Verguts & De Moor, 2005; see Nuerk and Willmes, 2005 for a review) and the overall magnitude of a two-digit number is achieved by integrating the single digits' of tens and units magnitudes into their respective position within the place-value structure of the Arabic number system (henceforth referred to as place-value integration, see also Poltrock & Schwartz, 1984 for an early inclusion of place-value knowledge into the decision process in number comparison). Thereby, the place-value structure is a necessary prerequisite for correctly representing multi-digit numbers. In the hybrid model (cf. Nuerk et al., 2001; Nuerk & Willmes, 2005) above described decomposed representation is accompanied by a holistic representation of the number's overall magnitude, meaning each number to be represented as an integrated entity of all its constituting digits without retaining the place-value structure. So, brought up by the observation of the compatibility effect (Nuerk et al., 2001) decomposed processing of tens and units in two-digit number comparison and on a broader level the possibility of an (additional) representation of number magnitude in terms of separate coding of each power level (cf. McCloskey, 1992) had to be considered as an alternative to the traditional assumption of an analogue holistic representation of all numbers disregarding the place-value structure of the Arabic number system (cf. Dehaene et al., 1990).

Nevertheless, although there is cumulating evidence for the notion of decomposed processing of multi-digit numbers from more and more numerical tasks (see below for a brief review), questions concerning representational universality / generality versus specificity are still under debate. Therefore, it was the aim of the current thesis to provide converging evidence for the existence and the importance of the place-value structure of the Arabic number system being retained in the human representation of number magnitude. This issue was pursued by evaluating effects possibly driven by separate representations of the individual power levels of multi-digit numbers in a variety of numerical tasks, using different methodologies and assessing diverse populations. However, before turning to a more explicit description of the present thesis recent evidence suggesting the place-value structure of the Arabic number system to be represented in the human understanding of number semantics will be reviewed briefly.

Further evidence for decomposed processing of tens and units

In recent years evidence for the existence of decomposed processing of number magnitude has accumulated. Extending the original results recent studies also using a number comparison paradigm revealed that the influence of the place-value structure of the Arabic number system on numerical cognition is generally language invariant (Moeller, Fischer, Nuerk, & Willmes, 2009a; Verguts & De Moor, 2005), although slightly moderated by characteristics of a language's number word system (Nuerk et al., 2005; Pixner, 2009), starts early in numerical development (Nuerk, Kaufmann, Zoppoth, & Willmes, 2004b; Pixner, Moeller, Zuber, & Nuerk, 2009), and is relatively independent from literacy (Wood, Nuerk, Freitas, Freitas, & Willmes, 2006a for data of Brazilian semi illiterates). Furthermore, Moeller et al., 2009a showed that separate processing of tens and units is performed in parallel rather than sequentially as suggested by Poltrock and Schwartz (1984). Moreover, Wood, Nuerk, and Willmes (2006b) were able to pinpoint the neural correlates of place-value integration (i.e., the compatibility effect) to the depths of the intraparietal sulcus (IPS, see also Goebel, Johansen-Berg, Behrens, & Rushworth, 2004 and Knops, Nuerk, Sparing, Foltys, & Willmes, 2006 for TMS data corroborating such a notion). Apart from this, effects of place-value integration were not limited to studies employing number comparison tasks. Nuerk, Geppert, van Herten, and Willmes (2002b) found evidence for a reliable influence of place-value integration on performance in a verification version of the number bisection task (NBT). Response latencies for evaluating whether a central number also represented the arithmetic mean between the two outer numbers of a triplet were significantly prolonged for triplets crossing a decade boundary as compared to triplets staying within the same decade, even when controlling for the overall range of the triplet (e.g., 25_28_31 vs. 23_26_29). Additionally, several studies investigating the development of numerical cognition found converging evidence that place-value understanding is reflected in children's transcoding performance (e.g., Pixner, 2009; Zuber, Pixner, Moeller, & Nuerk, 2009; see also Krinzinger et al., in press for language effects), develops differently in different cultures and number word systems (Miura & Okamoto, 1989; Miura et al., 1994; Pixner, 2009), and even influenced the development of spatial numerical representations as conceptualized by the mental number line (Moeller, Pixner, Kaufmann, & Nuerk, 2009b). Finally, processes of place-value integration have also been identified to determine performance in arithmetic tasks such as addition and multiplication. In addition the requirement of a carry is generally agreed to determine task difficulty (e.g., Kong et al., 2005; Deschuyteneer, de Rammelaere, & Fias, 2005). The execution of a carry in two-digit addition requires updating the decade digit of the result by the to-be-carried decade digit of the unit sum (i.e., 36 + 29 → 6 + 9 = 15 with 5 being the unit digit of the result and 30 + 20 + 10 = 60 with the last addend reflecting the to-be-carried decade digit of the unit sum, e.g., Klein et al., under revision) – a specific variant of place-value integration (analogue for the borrowing effect in subtraction). For multiplication the case is a little bit different. Generally, it has been observed for multiplication that responses are easier (i) for problems involving small as compared to large operands (e.g., 2 x 3 is easier than 8 x 7, e.g., Stazyk, Ashcraft, & Hamann, 1982), (ii) for problems involving 5 as an operand (e.g., 8 x 5 is easier than 8 x 4; Siegler, 1988) and (iii) for tie problems (e.g., 7 x 7, Lefevre et al., 1996). Additionally, Campbell & Graham (1985; see also Butterworth, Marchesini, & Girelli, 2003; Campell, 1994; 1997) found that the majority of multiplication errors are so-called operand errors in which the erroneous result represents the correct result of a neighbouring problem [i.e., in which one of the operands is mistakenly decreased or increased by 1, e.g., 4 x 6 = 28 which would be correct for 4 x (6 + 1)]. Evidence for decomposed processing of tens and units in multiplication comes from two observations. First, Verguts and Fias (2005a) were able to show in a computational modelling approach that problem size, five, and tie effect can be accounted for by a model assuming an initially decomposed representation of tens and units of the correct result. Second, Verguts and Fias (2005b; see also Domahs, Delazer, & Nuerk, 2006; Domahs et al., 2007) introduced the concept of unit-decade consistency to account for performance specificities in multiplication as described above. In this context consistency basically reflects the extent to which two neighbouring problems share the same decade or unit digit [e.g., 4 x 8 = 32; consistent probe: **3**6 = 4 x (8 + 1) vs. inconsistent probe: 28 = 4 x (8 – 1)]. The finding that in multiplication verification decade consistent, but nevertheless incorrect solutions probes, are harder to reject than decade inconsistent probes (Domahs et al., 2007, see Domahs et al., 2006; Verguts & Fias, 2005b for data from a production task again argues for a decomposed representation of two-digit numbers. In this vein, processing a decade digit identical to the decade digit of the correct result might drive the decision in one direction even when this decision bias is not confirmed by the processing of the unit digit, thereby suggesting separate processing of tens and units.

Taken together, recent empirical evidence (as outlined above) clearly suggests an important role of processes of place-value integration not only in very basic numerical task such as magnitude comparison but also in more complex tasks involving the coordination and integration of different numerical representations (e.g., magnitude, parity and fact knowledge in the NBT) and even in basic arithmetical operations. This indicates that place-value understanding is an important factor to correctly integrate the single digits constituting a given number into one comprehensive representation of this number. However, as already described above, the nature of the mental representation of multi-digit numbers is still debatable. Even though there is accumulating evidence corroborating the notion of decomposed processing of multi-digit number complying with the place-value structure of the Arabic number system representational aspects of place-value coding are still to be clarified. Even more particularly, questions concerning the importance of place-value understanding for the development of numerical / arithmetical capabilities are still untouched and remain to be to be investigated – as intended by the current thesis.

At the same time above arguments raise the question whether the place-value structure of the Arabic number system may be considered its own number representation apart from representational codes as introduced by Dehaene (1992; see also Dehaene & Cohen, 1995; 1997; Dehaene et al., 2003) in the Triple Code Model, for instance. This issue will be addressed in the next paragraph.

An independent representation of place-value structuring

Acknowledging the important role of an intact representation of the place-value structure of the Arabic number system in such a variety of numerical tasks Nuerk and colleagues (Nuerk, Graf, & Willmes, 2006) proposed place-value understanding to be an autonomous numerical representation (see also Moeller, Pixner, Klein, Cress, & Nuerk, 2009c). Based on the framework of the Triple Code Model (Dehaene & Cohen, 1995; 1997) Nuerk et al. (2006, see also Moeller et al., 2009c) derived six basic numerical representations: (i) a visual number form necessary to decode number symbols (ii) a semantic representation of number magnitude coding the quantity a number represents (iii) a verbal representation of numbers required for transcoding, additionally, it is assumed that numerical fact knowledge (e.g., multiplication facts) is represented in verbal format (iv) a spatial representation of number magnitude, i.e., a mental number line (v) conceptual, procedural, and strategic knowledge basically involved in performing arithmetic operations (e.g., 2 + 5 = 5 + 2 or 4 x 0 = 0)[1] as well as (vi) a representation of the place-value structure of the Arabic number system. Unlike other conceptualizations associating numerical cognition with an underlying general number sense (cf. Dehaene, 1997) or a number module (cf. Butterworth, 1999) the conceptualization by Nuerk and co-workers (2006) considers numerical cognition as an interplay of multiple basic numerical representations.

Each of these basic components contributes to mature numerical cognition in a very specific way. For instance, Holloway and Ansari (2009) observed that children who performed better in a number comparison task (indicating a more elaborate number magnitude representation) also scored higher in a standardized mathematics achievement test. Moreover, Dehaene and Cohen (1997) reported a double dissociation between the representations of number magnitude and multiplication fact knowledge in brain damaged patients indicating specific influences of either basic representation to numerical cognition (see also Delazer, Karner, Zamarian, Donnemiller, & Benke, 2006; Lemer, Dehaene, Spelke, & Cohen, 2003 for similar observations). Furthermore, Booth and Siegler (2008) were able to show that children

[1] Please note that this category is quite heterogeneous and can be further differentiated. For instance, conceptual knowledge such as 4 x 0 = 0 may reflect a different representation as compared to more procedural aspects such as knowing when a carry is needed in addition and how it is executed. Finally, strategic aspects may involve e.g., children's strategy to count on from the larger of two numbers that are to be added (i.e., 2 + 7 = 7 + 1 + 1).

whose spatial representation of number magnitude is more precise also achieved higher scores in a mathematics achievement test. Additionally, a better spatial representation of number magnitude was also predictive of children's learning of unfamiliar arithmetic problems.

In summary, it can be concluded that for most of above introduced basic numerical representations their importance for numerical cognition in general has already been subject to several studies. However, so far a comprehensive evaluation of the importance of processes of unit-decade integration into the place-value structure of the Arabic number system is still missing. The current thesis aimed at closing this gap by evaluating influences and implications of place-value integration for questions addressing the (i) nature of two-digit number magnitude representation, (i) the development of place-value representations, (iii) the neurocognitive underpinnings of unit-decade integration/place-value understanding, as well as (iv) a computational model replicating basic numerical effects. The motivation behind each of these issues will be introduced briefly.

Regarding the nature of two-digit number representation accumulating evidence suggests that decomposed processing of tens and units exists (see above). However, the generality of decomposed processing is still debatable. On the one hand, recent research (e.g., Ganor-Stern et al., 2009; Zhou et al., 2008) claimed that decomposed processing of tens and units in two-digit number comparison is limited to the case when the two to-be-compared numbers are presented simultaneously, i.e., externally when arguing with Zhang & Norman (1995; see also Zhang & Wang, 2005). On the other hand, so far, evidence for decomposed processing of two-digit number magnitude has only been reported for very basic numerical tasks such as number magnitude comparison (e.g., Nuerk et al., 2001; Moeller et al., 2009a; Ratinckx et al., 2006) and number bisection (Nuerk et al., 2002b) while evidence for decomposed processing of multi-digit numbers in more complex arithmetical tasks is still scarce (but see Domahs et al., 2006; 2007). Therefore, the general aim of the current thesis was to provide further evidence for the generality of a decomposed representation of two-digit number magnitude. On the level of basic research on the representation of two-digit number magnitude this goal was pursued by two studies. Based on the origins of the notion of decomposed processing of two-digit numbers Study 1 investigated whether there are indeed representational differences between internal and external magnitude representations as claimed by Zhang and Wang (2005) with only the external representation being decomposed complying with the place-value structure of the Arabic number system while the internal representation of number magnitude were holistic instead. Going beyond the case of such basic numerical tasks Study 2 was intended to offer new evidence for the generalizability of decomposed processing of tens and units to the case of mental arithmetic by evaluating the case of two-digit addition. However, when interested in the general validity of the account of decomposed processing of multi-digit number magnitude an important aspect refers to the influence of place-value properties on the development of numerical cognition which will be addressed in Studies 3 and 4. Study 3 follows up on a recent study by Moeller et al. (2009b) in which the authors presented an alternative account for the development of the spatial representation of number magnitude up to 100 suggesting two separate representations of single- and two-digit numbers; thereby considering place-value properties of the Arabic number system. In Study 3 of the current thesis the validity of their account is evaluated in a trans-lingual study also allowing for an evaluation of influences determined by differences in the number word systems of German and Italian on a non-verbal numeric task, i.e., the number line estimation task. Thus, apart from the obvious investigation of place-value influences in numerical development Study 3 also aimed at extending the notion of decomposed processing of tens and units to the case of the number line task.

From a developmental point of view it is interesting that none of the current models of numerical development (e.g., von Aster & Shalev, 2007; Butterworth, 2005; Wilson & Dehaene, 2007; Rubinsten & Henik, 2009) considers the role of place-value understanding, explicitly. Nevertheless, based on recent findings suggesting particular effects of place-value processing in children (e.g., Nuerk et al., 2004; Moeller et al., 2009c; Pixner et al., 2009) Study 4 specifically addressed the question of place-value understanding in numerical development in a longitudinal design. Therein, indices of place-value processing from a variety of basic numerical tasks such as magnitude comparison and transcoding were assessed in first grade and their predictive value for children's performance in an addition task administered two years later was inspected. To date, such longitudinal influences have only been investigated for the case of children's number magnitude representation (Holloway & Ansari, 2009; Halberda, Mazzocco, & Feigenson, 2008) but not for structural aspects such as place-value understanding. So, Study 4 of the current thesis aimed at replicating previous findings but for another basic numerical representation by combining specific aspects of prior studies (e.g., Zuber et al., 2009 on the influence of place-value information in transcoding; Pixner et al., 2009 on place-value integration in first-graders performing a two-digit magnitude comparison; Study 2 on decomposed processing in two-digit addition) to investigate the longitudinal development of numerical capabilities as determined by children's mastery of the place-value structure of the Arabic number system.

As the evidence corroborating the notion of decomposed processing of multi-digit number from classical behavioural and developmental studies accumulates, further neuro-psychological aspects regarding such processing are of particular interest. On the one hand, the question if and if so in what way place-value processing is affected by neuropsychological disorders such as hemi-spatial neglect arises. On the other hand, the question regarding possible neural correlates of place-value processing comes up. Both of these were addressed by Studies 5 and 6 of the current thesis. Against the background of the findings suggesting a disruption of the mental number line by hemi-spatial neglect (e.g., Zorzi, Priftis, & Umltà, 2002; Priftis, Zorzi, Meneghello, Marenzi, & Umltà, 2006) and the suggested existence of multiple number lines (i.e., one for each power level: units, tens, etc, see Nuerk et al., 2001; Nuerk & Willmes, 2005 for a more detailed discussion) the question comes up in how far hemi-spatial neglect affects place-value attributes of the to-be-processed numbers. In Study 5 the number bisection task was employed to evaluate the consequences of hemi-spatial neglect on the processing of place-value information. Thereby, using a task for which a crucial influence of place-value knowledge on task performance is known (see Nuerk et al., 2002) to assess place-value influences in neuropsychologically relevant group of participants to increase our knowledge on the generalizability of decomposed processing of two-digit numbers. In Study 6 the same task is again employed to evaluate possible neural correlates of place-value processing in particular. Wood and colleagues (2006b) were able to narrow down specific processes of place-value integration as reflected by the compatibility effect in two-digit magnitude comparison to a small area within the intraparietal sulcus. Thus, it was hypothesized that place-value processing as required in the number bisection task should lead to specific neural activation at a comparable site. Thereby, converging evidence for specific neural processing related to place-value representations would be provided from a different task, again corroborating the notion of decomposed processing of tens and units to be a quite general characteristic of multi-digit number processing.

Finally, the interpretation of empirical evidence should be guided by theoretical considerations, possibly formalized by a theoretical model. Nuerk and Willmes (2005, see also Nuerk et al., 2001 for first sketches and Knops, 2006 for further specifications) provided such a model of decomposed processing for the case of two-digit number comparison. Moreover, in recent years, a growing number of studies tried to implement the constraints of particular theoretical considerations on the mental representation of number magnitude into computational neural network models to evaluate differing notions regarding scaling properties of the mental number line, for instance (e.g., Zorzi & Butterworth, 1999; Dehaene & Changeaux, 1993; Grossberg & Repin, 2003; Verguts & Fias, 2004; 2008; Verguts, Fias, & Stevens, 2005). However, there is currently no computational model trying to realize a decomposed representation of two-digit number magnitude. Actually, most models did not even aim to account for two-digit number magnitude but were focused on single-digit or, at a maximum, teen numbers (but see Grossberg & Repin, 2003). Therefore, apart from promoting the generalizability of the notion of decomposed processing of two-digit numbers, an integral aim of the current thesis was to provide first computational evidence corroborating recent empirical results. In this vein, Study 7 pursued this goal by a direct comparison of three computational model each reflecting one of the current models of number magnitude representation (i.e., holistic, strictly decomposed, or hybrid, see above) by evaluating in how far modelled data replicated empirical effects such as the distance and unit-decade compatibility effect. Thereby, we wished to add a further line of evidence for the validity of the notion of decomposed processing of two-digit number magnitude the findings from classic behavioural (Studies 1 and 2), developmental (Studies 3 and 4), and neuro-psychological (Studies 5 and 6) experiments.

In the following section, the theoretical background and the intention of the studies pursuing above introduced issues will be discussed briefly to serve as an outline of what will be investigated why and in which way in each of the individual studies.

OVERVIEW OVER THE STUDIES INCORPORATED INTO THE PRESENT THESIS

As already mentioned above, the current thesis set off to investigate the importance of place-value understanding under different empirical aspects of two-digit number processing. The two studies of the second section follow up on basic research questions clarifying representational characteristics of two-digit numbers regarding place-value integration in (i) behavioural number magnitude comparison task (cf. Study 1) as well as (ii) in an eye-tracking experiment on addition verification (cf. Study 2). In the third section two studies evaluating (i) the influence of the place-value structure of the Arabic number system on the development of the spatial representation of number magnitude (cf. Study 3) as well as (ii) the longitudinal developmental trajectories of early place-value understanding for later arithmetic performance will be presented (cf. Study 4). Following on these developmental aspects section 4 addresses neuropsychological questions: (i) the neural correlates of place-value integration will be examined in a within task approach using functional magnetic resonance imaging (cf. Study 5) and (ii) the consequences of representational hemi-spatial neglect on place-value integration will be dealt with in patient study (cf. Study 6). In section 5 a computational model of two-digit number processing is introduced and validated by its performance in two-digit number magnitude comparison. Finally, section 6 shall provide a general discussion of the results of the individual studies together with a conclusive evaluation and further perspectives for the relevance and the investigation of the representation of the place-value structure as present in the Arabic number system.

Before presenting each of these studies the most important points reflecting the motivation and the rationale of these studies shall be considered in turn.

Section 2: On task general influences of the place-value structure of the Arabic number system in human number representation

1. **Study 1:** Internal number magnitude representation is not holistic, either.

   Regarding the way in which two-digit number magnitude is represented there is still an ongoing controversy among researchers. After the unit-decade compatibility effect (indicating decomposed processing of tens and units) has been published it was criticised to be just a perceptual artefact of the column-wise presentation of the two to-be-compared numbers one below the other. By replicating the compatibility effect with different display layouts (Nuerk et al, 2004) as well as for German number words (Nuerk et al., 2005) these notion could be refuted. Nevertheless, in 2005 Zhang and Wang presented evidence suggesting that decomposed processing of tens and units may depend on the presentational format employed (see also Zhou et al., 2008; Ganor-Stern et al., 2009). The authors found evidence for separate processing of tens and units when the standard to which a given probe number had to be compared and the probe itself were externally displayed at the same time. On the other hand, they did not observe any indication of decomposed processing when the probe had to be compared to an internally memorized standard. In the current study possible reasons why decomposed processing of tens and units may not have been detected in previous studies using a fixed internal standard are outlined. Additionally, an experiment is presented in which most of these confounds were controlled for.

2. **Study 2:** On the cognitive instantiation of the carry effect – Evidence from eye-tracking.

   Previously, the carry effect in addition was assumed to reflect the execution of asemantic procedural rules rather than involving specific magnitude processing. Such a view was corroborated by a variety of neuro-imaging studies in which processing carry addition problems as compared to non-carry problems was associated with increased frontal activation indicating increased demands on e.g., working memory (e.g., Ashcraft & Kirk, 2001; Fürst & Hitch, 2000), cognitive control (e.g., Miller, 2000) or . At the same time no specific increase in parietal activation around the IPS (known to subserve the processing of magnitude information) was found. However, recently, the data of Klein and Co-workers (under revision) suggested a special role of the unit digits of the summands in two-digit addition verification. In several regression analyses the authors could show that the carry effect did not seem to be purely categorical. Instead, the continuous predictor reflecting the sum of the unit digits of the summands turned out to be a reliable predictor of task performance. The specific influence of the unit digits only again indicated decomposed processing of two-digit numbers even in the case of basic arithmetic in the way that tens and units of two-digit numbers are seemingly processed separately rather than holistically as their influences on performance can be dissociated. However, the question remains whether it is the recognition of a carry that makes carry addition problems more difficult (only if the sum of the unit digits of the summands is equal or larger than 10 a carry is required) or rather the execution of the carry procedure involving processes of place-value, i.e., unit-decade integration as indicated by specific processing of either tens or units. The present study addressed this issue by recording participants' eye fixation behaviour in a two-digit addition verification paradigm. According to the eye-mind and the immediacy assumption on eye fixation behaviour above mentioned processes of recognizing vs. executing the carry should be distinguishable from each other. While the former should be associated with a specific increase of fixations on the unit digits of the summands for carry compared to non-carry addition problems, the latter should lead to particular increase of fixations on the decade digit of the result.

Section 3: On the importance of place-value integration for arithmetic development

1. **Study 3:** Language effects on children's non-verbal number line estimations.

   In a recent study Zuber et al. (2009) showed the importance of place-value understanding for transcoding numbers from one notation into another as well as the specific difficulties arising in a language (such as German) in which the order of tens and units is reversed in number words as compared to their symbolic instantiation when written in Arabic digits. Moreover, Moeller et al. (2009b) showed that in German speaking children place-value understanding is related to the development of the spatial representation of number magnitude: mastery of the place-value structure of the Arabic number system seems to be associated with the precision of children's estimates when asked to mark the position of a given number on a hypothetical number line. Children's estimation performance was accounted for best by a two-linear model suggesting two separate representations of single- and two-digit numbers indicating different processing of numbers which require or do not require place-value integration to correctly understand their overall magnitude, thereby, again highlighting the impact of place-value properties. However, from these studies it remains questionable whether this relation is specific for the German language (maybe driven by its inversion property) or rather culture invariant. This question was pursued in a cross cultural study replicating the results of Moeller et al. (2009b) in Italian speaking children who use a number word system without inversion and contrasting the results of German and Italian-speaking children in regard to the influence of place-value understanding on the results in a number line task.. This is particularly interesting as it would indicate influences of verbal number word representations on performance in a non-verbal numerical task and thus raising the question of the influence of verbal recoding on performance in non-verbal task.

2. **Study 4:** Early place-value understanding as a precursor for later arithmetic performance – a longitudinal study on numerical development.

In a cross-sectional approach Holloway and Ansari (2009) showed that an elaborate basic numerical representation of number magnitude as indicated by the performance in a number comparison task is associated with children's performance in a standardized mathematics achievement test. The better children performed in the comparison task the higher did they score in the achievement test (see Booth & Siegler, 2008 for a similar interrelation of the spatial representation of number magnitude and mathematics achievement). However, the possible importance of early place-value understanding for later numerical/mathematical development has not been evaluated yet. Based on the notion of multiple numerical representation to determine numerical cognition (see above) it is of interest on a theoretical level whether measures of early place-value understanding such as the transcoding performance (cf. Zuber et al., 2009) and more specifically the number of transcoding errors related to inversion in German-speaking children as well as the compatibility effect (cf. Pixner et al., 2009) is predictive of later arithmetic performance in general and for effects of place-value integration in later arithmetic (i.e., the carry effect in addition), in particular. The current study addressed this question by conducting a longitudinal investigation of whether and how transcoding and/or number comparison performance (in particular inversion errors and/or the compatibility effect) observed in first grade assessment are reliable predictors of addition performance in third grade and more particularly the carry effect (being driven specific and prominent processing of the unit digits indicating decomposed processing of two-digit number magnitude, see also Study 2) as obtained in third grade assessment.

Section 4: The neuro-cognitive underpinnings of place-value integration

1. **Study 5:** All for one but not one for all: How multiple number representations are recruited in one numerical task.

Recently, Goebel et al. (2004) were the first to dissociate the neural correlates of the representations of single- and two-digit numbers within the IPS suggesting separate representations for these two power levels. Further on, Wood and colleagues (2006b) investigated the neural underpinnings of the unit-decade compatibility effect to evaluate whether there is a specific cortex area (possibly within the IPS) subserving processes of unit-decade integration. The data by Wood and colleagues (2006b, see also Knops et al., 2006) were informative on this point and indicated that indeed neural activation in a circumscribed part of the IPS was determined by unit-decade compatibility. However, these results were obtained in a magnitude comparison task, assumed to primarily assess the representation of number magnitude. Therefore, it would have been interesting to evaluate the validity of these results in a more complex task known to recruit different numerical representation and requiring participants to integrate these representations to solve the task. The NBT fulfilled these criteria (cf. Nuerk et al., 2002b). In an fMRI using a verification version of the NBT we aimed at investigating the influence of processes of unit-decade integration, in this case the impact of decade crossings on task performance (e.g., 24_28_32 vs. 20_24_28; see Nuerk et al., 2002b).

2. **Study 6:** Impairments of the mental number line for two-digit numbers in neglect.

In their seminal paper Zorzi et al. (2002; see also Zorzi, Priftis, Meneghello, Marenzi, & Umiltà., 2006) observed that hemi-spatial neglect not only distorts the representation of intra- and extra-personal space but that this impairment generalizes to the spatial representation of number magnitude, thereby disrupting the mental number line. Interestingly, the spatial mental representation of number magnitude was affected by hemi-spatial neglect in the same way as mental representation of physical space. In a NBT with numbers up to 29 Zorzi and co-workers (2002) observed that patients' bisection errors mirrored the error pattern in physical line bisection: first, neglect patients misplaced the numerical mean of a given interval to the right (e.g., indicating 7 to be the midpoint between 1 and 9), second, the deviation between the correct and the indicated midpoint of a given interval increased as the range of the interval increased, and third, a cross-over effect was observed for very small interval ranges. As the NBT requires orienting and navigating along the mental number line we were interested whether such neglect effects could also be observed for the whole range of two-digit numbers. More particularly, we were interested whether hemi-spatial neglect had influences on the processing of place-value information as well. According to Nuerk and colleagues (2001, see also Nuerk and Willmes, 2005) there may not exist just one analoge mental number line but separate number lines for the different power levels of the Arabic number system (i.e., tens, units, etc.). Consequently, evaluation of the performance of neglect patients in a two-digit number bisection task may be informative for clarifying whether processing of structural aspects of the Arabic number system is also impaired in neglect patients. Therefore, the impact of neglect on the processing of decade crossings within the to-be-judged on triplets is of particular interest as in these triplets processing of place-value intergration is specifically relevant. When neglect patients would exhibit particular impairments in these triplets this would indicate that neglect not only hinders the processing of magnitude information but also of structural information when presented with two-digit numbers. In turn, this would imply that the mental number line incorporates such structural place-value information.

Section 5: A computational model of place-value integration in two-digit number comparison

Study 7: Two-digit number processing – holistic, decomposed or hybrid? A computational modelling approach

This last study is not an empirical study, but aimed at clarifying an issue for which there is currently no empirical resolution. At the moment there are three different conceptualizations of how number magnitude representation may be organized. The assumption of a holistic representation sees the magnitude of a number to be represented as a single integrated entity, thereby not retaining the place-value structure of the Arabic number system. Contrarily, the strictly decomposed view assumes that each power level of a given number (e.g., tens and units of two-digit numbers) is represented separately. Finally, the hybrid model of number magnitude representation is a combination of the former two models assuming both decomposed representations of the individual power levels as well as an overall holistic representation of a given number. In particular, the strictly decomposed and the hybrid model cannot be differentiated on the based of empirical data. Therefore, three network models corresponding to each of above mentioned representational models were programmed to evaluate which of these models accounts best for the accumulated empirical data. Particular interest was paid whether the models would replicate processes of unit-decade integration in number comparison, namely the compatibility effect. Furthermore, we aimed at investigating in how far the model data correspond to the empirical data for both comparisons to an external (e.g., Nuerk et al., 2001, Moeller et al., 2009a) as well as internal standard (cf. Moeller et al., 2009d, Study 1 of this thesis). Therefore, the data produced by either of the three models was contrasted to the empirical data of the studies by Moeller et al. (2009a, using a variable standard, i.e., singling out the larger number of a pair of numbers) and Moeller et al., 2009d, using an fixed standard, i.e., comparing a given number to an internally memorized standard).

Taken together, the current thesis aimed at investigating the influence of the place value structure of the Arabic number system on two-digit number processing in respect of four different aspects: (i) further clarification of basic representational characteristics of place value integration in magnitude comparison and addition (cf. Studies 1 and 2), (ii) investigating the developmental implications of early place value understanding (cf. Studies 3 and 4), (iii) evaluating the neuropsychological correlates of place value integration and its impairments (cf. Studies 5 and 6), (iv) introducing a computational model of two-digit number processing validated by empirical results pursuing basic representational questions (cf. Study 7). In the following, the individual studies, each addressing one of above mentioned issues, will be presented in turn.

OBJECTIVES BEYOND THE SCOPE OF THE INDIVIDUAL STUDIES

On a more general and abstract but less structuring level, the main aim of the current thesis was to provide a comprehensive evaluation of the impact of structural place-value attributes of the Arabic number system on numerical cognition. In the general discussion this goal will be pursued by considering and evaluating two different lines of evidence offered by the compilation of studies of the current thesis.

The first line of evidence to be followed will appraise the impact of the results of the current thesis as regards the generalizability of decomposed processing of two-digit number magnitude as general processing characteristic in multi-digit number processing. Therefore, particular interest was paid to new observations extend existing evidence [i] for the nature of the mental number line in terms of external vs. internal number magnitude representation and their impairments by hemi-spatial neglect (see Study 1 and 6), [ii] by evidence for decomposed processing from other basic numerical tasks such as the number line estimation task (see Study 3) as well as basic arithmetic procedures such as addition (see Studies 2 and 4), [iii] by evaluating longitudinal developmental influences of children's early place-value knowledge, which have been investigated exclusively for the case of number magnitude representations so far (see Studies 3 and 4), [iv] by investigating processing specificities of decomposed processing of two-digit numbers in respect of their neural correlates as well as its computational validity as compared to other models of number magnitude representation (see Studies 5 and 7).

The second line of evidence considered in the general discussion will evaluate in what way the results of the individual study not only provide evidence for a general processing principle driven by place-value properties of the Arabic number system. Instead, evidence on the relevance of individual (cf. Study 4) and cultural differences (cf. Study 3) in the influence of the place-value structure of the Arabic number system on numerical cognition and its determinants shall be reviewed. Moreover, as the evidence for individual and cultural differences as observed in the current thesis comes from developmental studies, the consequences and implications of these results for the case of adult number processing will be discussed and a possible implementation of place-value understanding into a current model of the development of numerical cognition (von Aster & Shalev, 2007) is suggested.

BIBLIOGRAPHY OF THE STUDIES INCORPORATED IN THE CURRENT STUDY

Section 2: On task general influences of the place-value structure of the Arabic number system in human number representation

Study 1: Moeller, K., Nuerk, H.-C., & Willmes, K. (2009). Internal number magnitude representation is not holistic, either. *The European Journal of Cognitive Psychology, 21,* 672.

Study 2: Moeller, K., Klein, E., & Nuerk, H.-C. (submitted). On the cognitive instantiation of the carry effect – Evidence from eye-tracking.

Section 3: On the importance of place-value integration for arithmetic development

Study 3: Helmreich, I., Zuber, J., Pixner, S., Kaufmann, L., Nuerk, H.-C., & Moeller, K. (accepted). Language effects on children's mental number line: How cross-cultural differences in number word systems affect spatial mappings of numbers in a non-verbal task. *Journal of Cross-Cultural Psychology.*

Study 4: Moeller, K., Zuber, J., Pixner, S., Kaufmann, L., & Nuerk, H.-C. (submitted). Early place-value understanding as a precursor for later arithmetic performance – a longitudinal study on numerical development.

Section 4: The neuro-cognitive underpinnings of place-value integration

Study 5: Wood, G., Nuerk, H.-C., Moeller, K., Geppert, B., Schnitker, R., Weber, J., & Willmes, K. (2008). All for one but not one for all: How multiple number representations are recruited in one numerical task. *Brain Research, 1187,* 154-166.

Study 6: Hoeckner, S., Moeller, K., Zauner, H., Wood, G., Haider, C., Gaßner, A., & Nuerk, H.-C. (2008). Impairments of the mental number line for two-digit numbers in neglect. *Cortex, 44,* 429-438.

Section 5: A computational model of place-value integration in two-digit number comparison

Study 7: Moeller, K., Huber, S., Nuerk, H.-C, & Willmes, K. (in preparation). Two-digit number processing – holistic, decomposed or hybrid? A computational modelling approach

Section 2

Task invariant influences of the place-value structure of the Arabic umber system on numerical cognition

Study 1

Internal number magnitude representation is not holistic, either

ABSTRACT

Over the last years, evidence accumulated that the magnitude of two-digit numbers is not only represented as one holistic entity, but also decomposed for tens and units. Recently, Zhang and Wang (2005) suggested such separate processing may be due to the presence of external representations of numbers while holistic processing became more likely when one of the to be compared numbers was already internalized. The latter conclusion essentially rested on unit-based null effects. However, Nuerk and Willmes (2005) argued that unfavourable stimulus selection may provoke such null effects and misleading conclusions consequently. Therefore, we tested the conclusion of Zhang and Wang for internal standards with a modified stimulus set. We observed reliable unit-based effects in all conditions contradicting the holistic model. Thus, decomposed representations of tens and units can also be demonstrated for internal standards. We conclude that decomposed magnitude processing of multi-digit numbers does not rely on external representations. Rather, even when two-digit numbers are internalized, the magnitudes of tens and units seem to be (also) represented separately.

INTRODUCTION

Number magnitude has long been assumed to be represented holistically along a mental number line, even for multi-digit Arabic numbers (Dehaene, Dupoux, & Mehler, 1990). In recent years, evidence has accumulated that questions the assumption of purely holistic two-digit number processing in different tasks such as magnitude comparison (Korvorst & Damian, in press; Nuerk, Weger, & Willmes, 2001; Verguts & de Moor, 2005, Wood, Mahr, & Nuerk, 2005), multiplication (Domahs, Delazer & Nuerk, 2006; Verguts & Fias, 2005), addition (Deschuyteneer, De Rammelaere, & Fias, 2005; Kong, Wang, Kwong, Vangel, Chua, & Gollub, 2005) or the number bisection task (Nuerk, Geppert, van Herten, & Willmes, 2002; Wood et al., 2008). In sum, there is ample evidence now that multi-digit numbers can be processed in a decomposed fashion (for a review, see Nuerk & Willmes, 2005). Therewith, the debate has turned to the question under which conditions decomposed processing occurs.

At the core of this debate is the unit-decade compatibility effect (Nuerk, Weger, & Willmes, 2001; 2004; 2005). In two-digit number magnitude comparison compatible number pairs for which separate comparisons of tens and units bias the response in a similar direction (e.g. 42_57, $4 < 5$ and $2 < 7$) are evaluated faster than number pairs resulting in incompatible decision biases (47_62, $4 < 6$, but $7 > 2$). As overall distance was matched between compatible and incompatible pairs, the assumption of holistic number magnitude representation cannot account for the compatibility effect. Therefore, Nuerk et al. (2001) have suggested a hybrid model of two-digit number processing that assumes collateral mental number lines for tens and units in addition to a (holistic) mental number line representing overall number magnitude (see also Nuerk & Willmes, 2005, but see Verguts & de Moor, 2005).

Zhang and Wang (2005) argued that such separate representations of tens and units may be only or at least particularly valid for the external representation of two-digit numbers.

The authors used a magnitude comparison task with standard numbers 55 and 65. In an external representation condition, the standard was presented together with a probe whereas in the internal representation condition, the probe was presented alone and had to be compared to an internal standard presented 3 seconds before. For the external standard condition, Zhang and Wang observed reliable influences of the unit digits' magnitude on task performance in the external condition: (i) discontinuity effects (i.e. the RT difference between 69 and 70 differed significantly from the difference between 68 and 69) at decade boundaries or (ii) even a reversed distance effect (i.e. at the boundary of 30s and 40s RTs to probes 36 through 39 were on average slower than those to probes 41 through 44 even though overall distance is larger for the former probes). Quite to the contrary, the authors reported mixed results when the standard was internal. While they observed unit-based effects in a two-way (decades × units) ANOVA, they completely failed to find any discontinuity or reversed distance effects. Thus, Zhang and Wang concluded that the influence of the unit digit in number comparison is determined by the representational format. Only when both numbers (i.e. standard and probe) are represented externally, decomposed processing of tens and units is assumed.

However, we do not think that these null unit-effects warrant conclusions that decomposed representations have to rely on external representations of numbers. We have pointed out previously why null unit-effects could be observed (Nuerk & Willmes, 2005). As this claim has guided the design of the current study a brief sketch will be given on how such null unit effects come about.

How to observe unit-based null effects in two-digit number comparison

1. Use of small unit distances in the stimulus set

We observed repeatedly that the compatibility effect is more pronounced when unit distances are large and possibly not even present when unit distances are small (Nuerk et al., 2001; 2005; Nuerk, Kaufmann, Zoppoth, & Willmes, 2004). When using a standard with the unit 5 (e.g. 65), unit distances are limited to a maximum of 4 (when excluding multiples of ten). For such rather small unit differences, the unit-based compatibility effect is weakest and possibly not significant. Therefore, if one intends to find null unit-effects, standards like 65 or - for the comparison of two numbers - only small unit distances should be presented.

2. Use of no or only a few within-decade trials

We observed that the compatibility effect is larger when within-decade trials (e.g., 53_57) are included in the stimulus set (Knops, 2006; Nuerk & Willmes, 2005, for a review). So, when units are irrelevant in all or the vast majority (e.g. only 15.5 % within-decade trials in Dehaene et al., 1990) of trials, unit-effects may be smaller as attention may be focussed on the tens, as these are decisive in all or almost all trials. Consequently, such an attentional bias towards the tens' digit(s) leads to a weakening of possible unit-based effects. Therefore, if one wishes to observe null unit-effects using exclusively or at least a majority of between-decade trials would be a good advice.

3. Ignore congruity effects in magnitude comparison

We observed automatic within-number comparison of tens and units (e.g. for 62, 6 < 2; Wood et al., 2005) to influence response latencies when this respective number had to be compared to another number presented 100 ms later (e.g. 62_47). In this example the result of comparing tens and units within 62 is congruent with the overall decision: the decade digit of 62 is larger than its unit digit (i.e. 6 > 2) and is also larger than the decade digit of 47 (i.e. 6 > 4). Hence, the task-relevant decision (i.e. 62 > 47) is facilitated. Contrarily, respective but incongruent comparisons as for the number pair 42_57, i.e. the decade digit of 57 is smaller than its corresponding unit digit (i.e. 5 < 7) but larger than the decade digit of 42 (i.e. 5 > 4), inhibit the relevant decision (i.e. 42 < 57).

Such congruency is important as most compatible trials are incongruent whereas most incompatible trials are congruent (see Wood et al., 2005; Nuerk & Willmes, 2005). Therefore, when using a fixed internal standard the congruency effect works against the unit-based compatibility effect and helps to produce unit-based null effects.

Objectives

We have briefly outlined how unit-based null effects can most likely be produced. However, there is the problem that such null effects observed under the conditions stated above are not conclusive. If one does not find unit-based effects under unfavourable conditions, this does not mean that unit-based effects do not exist at all. Consequently, the absence of unit-based effects under unfavourable conditions does not imply holistic processing, especially not when unit-based effects can be obtained under more favourable conditions.

The current study aimed to evaluate the conclusion of a holistic internal representation of two-digit numbers as drawn by Zhang and Wang (2005) under more favourable conditions. In their study the authors did not control for the above three unfavourable conditions: (i) Only small unit distances were used as the standards were 55 and 65. (ii) The unit digits were largely irrelevant in their study as only 13 % of the trials required within-decade comparisons. (iii) In the internal representation condition, only one two-digit number was presented so that the congruency effect worked against unit-based effects.

Based on these considerations the objective of the study is straightforward: Do the results of Zhang and Wang (2005) indicating a holistic internal representation of two-digit numbers hold even under conditions more favourable to detect unit-based effects indicative of decomposed internal representations of tens and units? To directly test this, the experimental design of Zhang and Wang was changed in two essential aspects: (i) The standards 53 and 57 were chosen to allow for larger unit distances. (ii) 50% within-decade trials were used to preclude any attentional bias towards the tens, making decade and unit digits equally relevant. If two-digit numbers were represented internally in a holistic fashion, these changes should not matter. However, if numbers were represented internally in a decomposed fashion, these changes should be relevant and should produce significant unit-based effects, in particular the compatibility effect. Note that this prediction is made although we could do nothing about the third problem as the congruency effect is inevitably confounded with the examination of internal representations.

METHOD

Participants: 19 students of the University of Salzburg (11 female) participated in the experiment. Mean age was 21.8 years (SD: 2.4 years; range: 18 - 27 years). All participants reported normal or corrected to normal vision and were right-handed.

Stimuli: In the current experiment two-digit numbers ranging from 31 to 79 had to be compared to the two standard numbers 53 and 57. The stimulus set did not contain multipliers of ten and - depending on the standard employed - either the number 53 or 57 was omitted. Probe numbers smaller or larger than the standard were balanced in frequency of occurrence as was the number of between- (e.g. 53_45) and within-decade trials (e.g. 53_58). This latter requirement was met by selective repetitions of within-decade stimuli. Altogether, 72 probe numbers had to be compared to either standard. 36 probe numbers were between-decade stimuli, 18 smaller (31 - 49 excluding 40) and 18 larger than the standard (61 - 79 excluding 70). Additionally, 36 within-decade stimuli were presented, again 18 smaller and 18 larger than the standard. To attain this balanced stimulus set, for the standard 57, each number from 51 - 56 was presented 3 times, and 58 and 59 both 9 times, and accordingly we included 9 times both 51 and 52 and 3 times 54 - 59 for the standard 53.

All probes were shown in Arabic notation using the “Times New Roman” font (size 24) and were presented at the centre of a 17‘ monitor.

Procedure: After ten randomly chosen practice trials the first of 2 runs with 10 blocks each was started. Each block consisted of all 72 probe numbers for one of the two standards. Half of the participants started with ten blocks of comparing the probe to the internal standard 53 while the other half started with the internal standard 57[1]. Trial order was randomized for each participant separately. Viewing distance was approximately 60 cm.

In each trial, a fixation mark located at the centre of the screen was presented for 500 ms. Then the probe number appeared and remained on the screen until a response was given. Participants had to press the up-arrow key of a standard keyboard with their right index finger when the probe was larger than the standard and the down-arrow key with their left index finger when the probe was smaller than the standard.

Mode of analysis: For between-decade trials the factors *standard number* (53 vs. 57), *unit-decade compatibility* (compatible vs. incompatible, e.g. 57_43 vs. 53_37), and *unit distance* (small: 1 – 3 vs. large: 4 – 6) were discerned. Please note that no compatible/incompatible distinction is possible for probes deviating from the standard by a multiple of ten. Therefore, these probes were not considered in the analysis. Based on this classification, overall distance would have been larger for compatible than for incompatible number pairs making it impossible to distinguish holistic from decomposed accounts. Therefore, 6 selectively chosen compatible as well as incompatible probes were excluded based on the constraint of balancing overall distance for compatible and incompatible trials. Note that matched overall distance can only be achieved by excluding specific, not randomly chosen probes (see Appendix A for details). However, it is impossible to match both overall distance and problem size (½ * (standard + probe)). Therefore, the latter was incorporated in the analyses as a covariate. As problem size can only be computed for trials but not

[1] Please note that this procedure is even stricter than in Zhang and Wang's (2005) study as it leaves much more time to internalize standards. In Zhang and Wang, standards could be internalized for 3 seconds for one comparison, while in our study the standard could be internalized for 720 trials being more comparable to the procedure used by Dehaene and colleagues (1990).

participants for between-decade trials an ANCOVA was conducted over all included items incorporating above mentioned factors. Additionally, a stepwise regression analysis was run over all included items including the following predictors: (i) overall absolute distance, (ii) logarithm of the absolute distance, (iii) distance of the logarithmic magnitudes, (iv) decade distance, (v) logarithmic decade distance, (vi) problem size, (vii) logarithmic problem size, (viii) unit distance ranging from -6 for e.g. 71_57 to +6 for e.g. 53_69, (ix) absolute unit distance and (x) compatibility.

As within-decade trials were only included to preclude participants from focussing their attention exclusively on the decade digit, these were treated as filler items and no further analysis was carried out.

RESULTS

Overall mean error rate was very low (standard 53: 3.0 %, standard 57: 3.5 %) and will not be considered here in greater detail. However, in line with earlier findings (e.g. Nuerk et al., 2001) a reliable compatibility effect was present for error rates, too [$F(1, 18) = 7.06$, $p < .05$, compatible: 1.7 % vs. incompatible: 3.0 % errors].

ANCOVA: Although the incorporated covariate, i.e. problem size, accounted for a significant part of the variance [$F(1, 39) = 5.13$, $p < .05$], a reliable compatibility effect [$F(1, 39) = 8.22$, $p < .01$] was observed in the ANCOVA: compatible trials (492 ms) were followed by shorter response latencies than incompatible trials (504 ms)[2]. Neither the main effect of standard number [$F(1, 39) < 1$] or unit distance [$F(1, 39) < 1$] nor any interaction [all $F < 1.47$, all $p > .24$] was significant. However, when differentially evaluating the compatibility effect for small and large unit distances, separate ANCOVAs with problem size as a covariate showed that the overall main effect of compatibility was mainly driven by a reliable

[2] Note that the compatibility effect was also reliable when running the analysis over participants instead [12 ms, $F(1, 18) = 8.04$, $p < .01$] indicating that it was not moderated by problem size differences.

compatibility effect for large unit distances [16 ms, $F(1, 11) = 13.07$, $p < .01$, see Figure 1], while there was no significant compatibility effect for small unit distances [7 ms, $F(1, 27) = 2.34$, $p = .14$], suggesting particular influence of large unit-distances.

Regression: The regression analysis on mean RT corroborated the ANCOVA results. The multiple correlation was highly predictive ($R = .717$, adjusted $R^2 = .488$, $p < .001$) by incorporating only two significant predictors: Response latencies became faster with increasing distance between the logarithms of standard and probe ($\beta = - .599$) and increasing (compatible) unit distance ($\beta = - .413$). As unit distance ranges from -6 in incompatible to +6 in compatible trials this result indexed that the compatibility effect is not categorical, but increases as incompatible unit distance increases. Note that within-decade probes could not be included in the regression analyses as the (positive or negative) unit distance is defined with regard to the compatibility of the decade digits' comparison, for which the decade digits have to be different.

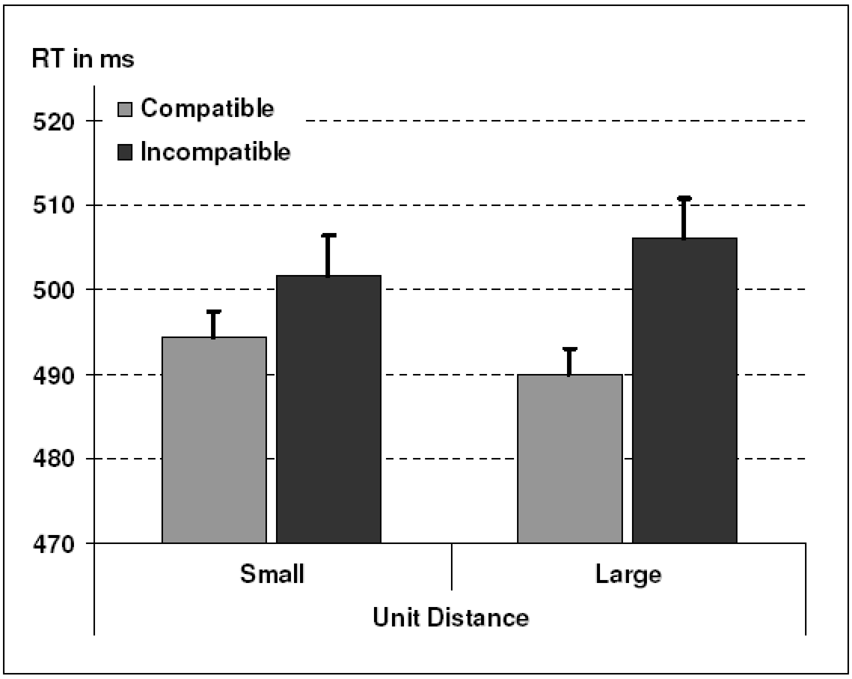


Figure 1: *Means of response latencies for the respective stimulus groups. Error bars indicate the standard error of the mean (SEM).*

DISCUSSION

Internal or external decomposed representations?

The objective of the present study was straightforward: Does the assumption of a strictly holistic internal representation of two-digit numbers hold under conditions more favourable to detect unit-based effects? The current data indicate that the answer should be no. When allowing for unit distances larger than 4 and employing a balanced number of between- and within-decade trials, a reliable unit-decade compatibility effect was observed. Such a compatibility effect in an internal condition cannot be explained by a holistic representation, but is consistent with an independent contribution of the unit digit's magnitude.

In line with the hypothesis, the changes introduced in the experimental design did help to observe unit-based effects: (i) besides the significant unit-decade compatibility effect, the present data also showed that this effect was mainly driven by a compatibility effect for large unit distance (i.e. 4 - 6). This not only corroborated the objections against the use of a standard involving the unit digit 5 as raised in the introduction, but also indicates that the compatibility effect is not a simple attentional congruity effect. It rather reflects processing of the unit digit's magnitude and thus points to the importance of the units in two-digit number comparison. (ii) In accordance with the findings of Knops (2006) the current results imply that the use of a considerable proportion of within-decade trials in which comparing the unit digits is decisive (e.g. 53_58) seems to be necessary to prevent participants from exclusively focusing their attention on the decade digits and therefore to reduce the possibility to obtain unit-based effects. Finally, the significant compatibility effect is particularly relevant as it emerges even against the impact of the congruency effect described in the introduction which is inevitably confounded with the use of an internally represented standard.

Although reliable, the compatibility effect in the present study was smaller than that reported in previous studies (12 ms vs. 31 ms in Nuerk et al., 2001; 29 ms in Nuerk et al., 2004). So, increasing the maximum unit distance from 4 (e.g. Zhang & Wang, 2005) to 6 contributed to the observation of the compatibility effect, but it still seems to be the largest possible unit distances (i.e. 7 and 8) that particularly enhance the compatibility effect. Nuerk and Willmes (2005) had hypothesized that such limited unit distance may account for the lack of unit-based effects in comparisons to a standard involving the unit digit 5. However, the current study is the first in which this hypothesis was directly tested.

However, even when these results are hard to reconcile with the notion of two-digit numbers to be exclusively represented as integrated entities, this does not preclude holistic processing of two-digit numbers entirely. Decomposed representations of the magnitudes of tens and units as indicated by the compatibility effect may exist in addition to an (approximate) holistic representation of their integrated magnitude (see also Liu, Wang, Corbly, Zhang, & Joseph, 2006 for neuro-imaging data corroborating this notion). Such a view has been expressed explicitly in the hybrid model of two-digit number representation (Nuerk & Willmes, 2005). The authors suggest that these three magnitude representations are activated in parallel and influence each other. In particular, an excitatory influence facilitating the correct response is assumed for the case of compatible separate comparisons of tens and units (e.g. 42_57, 4 < 5 and 2 < 7). In contrast, when separately comparing tens and units yields incompatible decision biases (e.g. 47_62, 4 < 6 but 7 > 2) the correct response is inhibited. The regression analysis in this study corroborates earlier analyses in that both predictors related to holistic processing (i.e. distance between the logarithms of standard and probe) and of decomposed processing (i.e. unit-distance) were entered into the model. However, it must be noted that as for the Nuerk et al. (2001) data (see Verguts & De Moor, 2005, for a reanalysis) the regression model does not get much worse if decade distance is forced to be a predictor in the regression analysis (adjusted R^2 = .486 compared to .488 in the original analysis). So, although we personally favour a hybrid processing model, a fully decomposed model cannot be rejected on the basis of these data.

The debate under which conditions holistic two-digit number processing dominates or replaces decomposed processing of tens and units continues (Zhou, Chen, Chen, & Dong, in press; see Nuerk & Willmes, 2005 for a review). The current data have a clear message concerning this controversy: decomposed processing does not rely on *external* representations, but can also be demonstrated even when *internal* representations are involved. This suggests that two-digit numbers can be represented in a decomposed fashion irrespective of their representational format (i.e. external or internal). And actually, this finding has an everyday implication as whenever we judge whether a price is expensive or not, we compare this price to an internal standard which deems us (approximately) appropriate. However, a possible limitation of this interpretation may be that both Zhang and Wang (2005) and Zhou et al. (in press) report unit-based null effects when the to be compared numbers were presented serially rather than simultaneously. Therefore, the particular influence of this distinction cannot be answered by the current results but has to be addressed in future research. Nevertheless, what the current results do indicate is that the decomposition of two-digit numbers is not limited to the external representational format but generalizes to the internal representation as well.

A word on the logarithmic fitting of overall distance

Previously, (e.g. Dehaene et al., 1990) logarithmic fittings of the distance effects for two-digit numbers have been used to indicate holistic rather than decomposed processing. It was argued before (Nuerk & Willmes, 2005, see also Verguts & de Moor, 2005, for similar arguments) that a good fit of overall logarithmic distance can be produced by decomposed as well as by holistic models. In the present study, some parameters have been changed as compared to previous studies: different standards (53 and 57) allowing for larger unit distances were used together with 50 % of within-decade trials as compared to a much lower proportions in earlier studies. Therefore, it might be an interesting question whether these changes also affect the goodness of logarithmic fitting.

In Figure 2, the results of fitting this logarithmic predictor can be found. Generally, the logarithmic fitting of overall distance is still good for the present design. However, it is influenced by the standard used. For the standard 53 fitting is better for numbers larger than 53 (see Figure 2, Panel A), whereas for the standard 57 fitting is better for numbers smaller than 57 (see Figure 2, Panel B). The reasons for this asymmetry might be twofold: (i) In the conditions showing better fits, there were more (i.e. 6) relatively slower within-decade numbers. Thus, the larger within-decade distance effect might have determined the logarithmic slope close to the standard. Contrarily, there were only few (i.e. 2) within-decade numbers in the conditions showing worse fits, which may not be favourable to lead to a pronounced within-decade distance effect. Additionally, these 2 within-decade numbers were repeated quite often which may have affected their RT to a certain extent[3]. (ii) The logarithmic fit was worse in the conditions with more incompatible trials. Considering the standard 53, there are 6 compatible trials per decade for probes larger than the standard (e.g. 53_67) but only 2 incompatible ones (e.g. 53_71) whereas this pattern is reversed for probes smaller than the standard (e.g. compatible: 53_41; incompatible: 53_37). For the standard 57 the compatible-incompatible pattern is reversed. The worse fit in these conditions may indicate that enhanced decomposed processing of tens and, in particular, units is required here to overcome unit-decade incompatibility. In turn, such an increase in decomposed processing may reduce the influence of the holistic representation as reflected in a worse fit.

[3] Note that the compatibility effect in this study was based on between-decade stimuli which were all presented equally often only once per block.

All in all, good logarithmic fits were observed in the current study underlining that even in the presence of decomposed processing, logarithmic overall distance can be an additional good predictor of performance (Nuerk & Willmes, 2005). However, the data also indicate that the steeper slope of the within-decade distance effects based on sufficient within-decade numbers as well as a balanced number of compatible and incompatible trials may be important for good logarithmic fits.

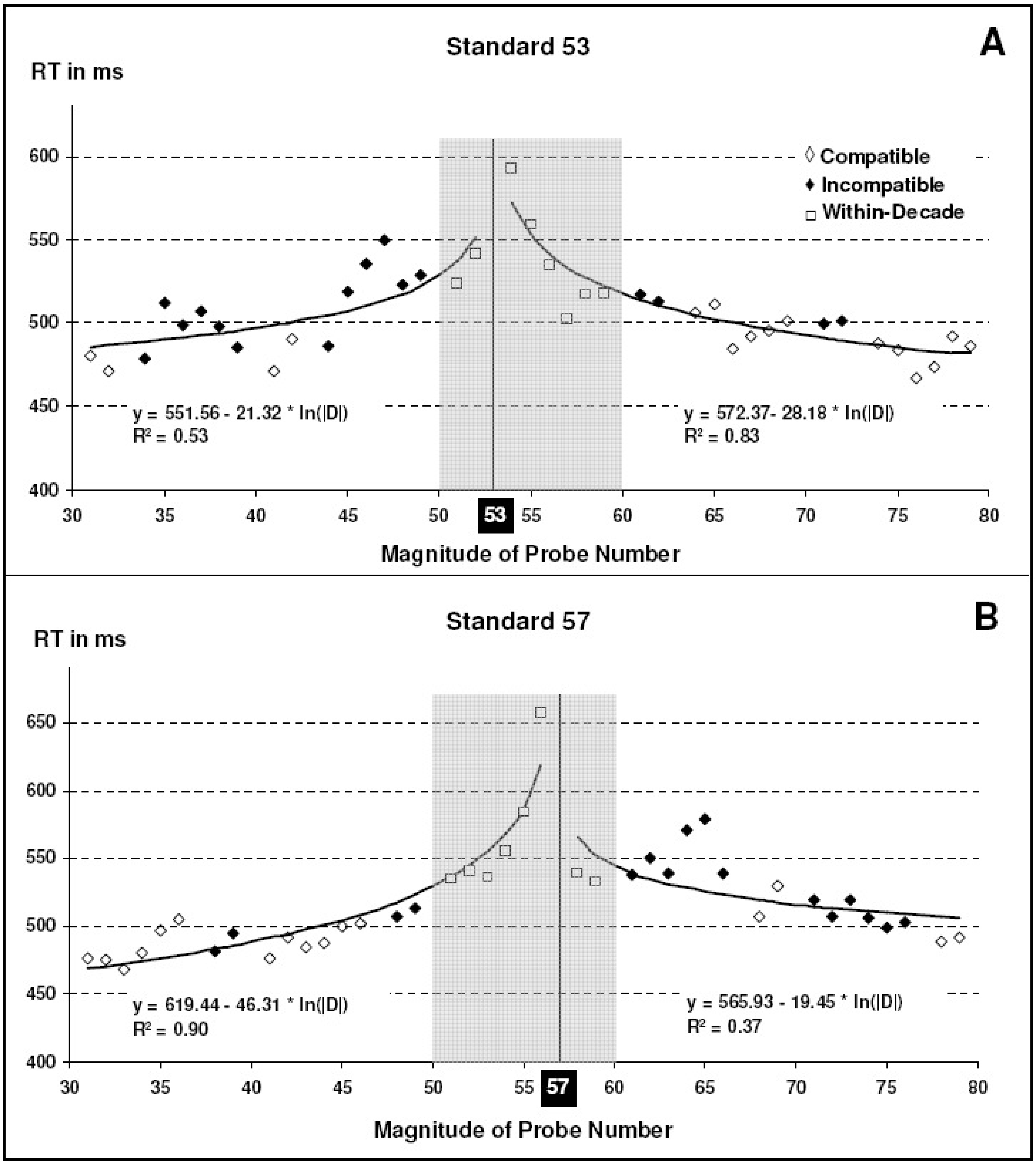


Figure 2: *Results of logarithmic fitting (regression equation and R^2) depicted separately for standard 53 (Panel A) and standard 57 (Panel B).*

CONCLUSIONS

In summary, the conclusions to be drawn are twofold: first, these data present evidence that the representation of two-digit numbers does not generally depend on representational format (i.e. internal vs. external) as proposed by Zhang and Wang (2005). If the internal representation of two-digit numbers were holistic, allowing for larger unit distances and balancing the number of within-decade trials should not have changed the results. However, as these changes mattered and produced a significant compatibility effect, internal two-digit number representation can hardly be considered to be exclusively holistic. Rather, these data suggest two-digit numbers to be represented in a decomposed fashion independent of an internal or external representational format for the comparison standard.

Second, these data illustrate that a careful choice of stimuli and experimental conditions is necessary in investigations of multi-digit number processing. Two modifications are considered particularly important in this study: (i) the use of large unit-distances for examining compatibility effects and (ii) the use of within-decade stimuli to prevent attentional biases. The data suggest that these factors should also be considered in future studies investigating the nature of multi-digit number processing under different conditions.

APPENDIX A

Overall distance and problem size of the respective stimulus groups

Standard	53				57			
Compatibility	compatible		incompatible		compatible		incompatible	
Unit Distance	small	large	small	large	small	large	small	large
Before adjustment								
Overall distance	16.8	20.0	13.2	10.0	16.8	20.0	13.2	10.0
Log overall distance	1.23	1.30	1.12	1.00	1.23	1.30	1.12	1.00
Problem size	54.8	63.0	51.8	48.0	55.2	47.0	58.2	62.0
After adjustment								
Overall distance	14.6	15.0	15.4	15.0	14.6	15.0	15.4	15.0
Log overall distance	1.16	1.18	1.19	1.18	1.16	1.18	1.19	1.18
Problem size	50.8	60.5	53.0	45.5	59.1	49.5	57.0	64.5

Adjustment comprised exclusion of selective trials to match overall distance between all stimulus groups. Such an adjustment is important because otherwise compatibility effects and distance effects are confounded and holistic processing cannot be distinguished from decomposed processing. Due to this adjustment the incompatible probes 44, 45, 46, 47, 48, 49 as well as the compatible probes 74, 75, 76, 77, 78, 79 were excluded in the standard 53 condition. Probes 31, 32, 33, 34, 35, 36 (compatible) and 61, 62, 63, 64, 65, 66 (incompatible) were omitted from the analysis for standard 57. Please note that the role of problems size was addressed in the ANCOVA and in the regression analysis.

<u>Study 2</u>

On the cognitive instantiation of the carry effect in addition

–

Evidence from eye-tracking

ABSTRACT

Recent research indicated that processes of unit-decade integration pose particular difficulty on multi-digit addition. In fact, longer response latencies as well as higher error rates have been observed for addition problems requiring a carry operation (e.g., 18 + 27) compared to problems not requiring a carry (e.g., 13 + 32). However, the cognitive instantiation of this carry effect remained unknown. In the current study this question was pursued by tracking participants' eye-movements during addition problem verification. Analyses of the eye-fixation data suggested a differential influence of decade and unit digits on mental addition: While the necessity of a carry operation specifically increased total reading times on the unit digits of the summands, total reading times were particularly increased on the decade digit of the result. This indicated that both recognizing the requirement of a carry (by means of calculating the sum of the unit digits of the summands) as well as its completion on the decade digit of the result determine the difficulty of carry addition problems. On a more general level, this study shows how the nature of numerical-cognitive processes can be further differentiated by the evaluation of eye movement measures.

INTRODUCTION

One of the probably most robust findings in multi-digit addition is the very strong influence of the requirement of a carry operation on task performance. Whenever a carry is required, both response latencies and error rates increase considerably (Deschuyteneer, De Rammelaere, & Fias, 2005; Fürst & Hitch, 2000; Imbo, Vandierendonck, & De Rammelaere, 2007; Klein, Nuerk, Wood, Knops, & Willmes, 2009; Kong et al., 2005). Generally, whether or not a carry operation is needed is determined by the summands of the addition problem: Whenever the sum of the unit digits of the summands is 10 or larger a carry is necessary to compute the correct result (e.g., 7 + 8 = 15 in 47 + 28), whereas no carry is needed whenever the sum of the units is smaller than 10 (e.g., 2 + 3 = 5 in 52 + 23). For the above example 47 + 28, the carry operation is executed by adding 1 (representing the decade digit of the unit sum) to the sum of the decade digits of the summands. In this case, the sum of the unit digits is 15, so the unit digit of the result (i.e., 75) is 5 and the decade digit of the result is derived by updating the sum of the decade digits of the summands by 1 (i.e., 4 + 2 + 1 = 7). However, not only the mere requirement of a carry operation influences performance. Rather, Imbo and colleagues (2007) showed that both response latencies and error rates also increased with the number of carries required in one addition problem (e.g., 81 + 56 = 137 vs. 59 + 78 = 137) as well as the value of a carry (e.g., in 24 + 18 + 29 unit sum equals 21 with the carry being 2).

Taken together, it is established that carry addition problems are more difficult than non-carry problems and that this difficulty further increases with the number as well as the value of carries required in a problem. However, our knowledge on the specificities determining this higher difficulty of carry addition problems is still patchy. In the following, different aspects proposed to account for the difficulty of carry addition problems shall be reviewed briefly.

Sources of difficulty in carry addition problems

On a theoretical/conceptual level difficulties arising with the requirement of at least one carry have been suggested to originate from different sources. On the one hand, Kalaman and LeFevre (2007) argued that the carry effect represents an increased demand on working memory resources for e.g., keeping track of intermediate results (see also Imbo et al., 2007; Kazui, Kitagaki, & Mori, 2000; Zago et al., 2001). On the other hand, Nuerk, Graf and Willmes (2006) associated the increased difficulty of carry addition problems with a higher workload for the correct classification (i.e., units, tens, etc.) and manipulation of digits within the base-10 structure of the Arabic number system (i.e., carrying from one position to another). Finally, Green, Lemaire, and Dufau (2007) suggested that the carry effect may also reflect processes necessary to adapt solution strategies to the actual problem (see also Torbeyns, Verschavel, & Ghesquiere, 2002). In sum, these data suggest that there seem to be different sources of difficulty for the carry operation.

At this point, evaluating eye fixation behaviour may supplement the interpretation of response latencies and error rates when investigating the temporal dynamics of number processing (cf. Brysbaert, 1995). According to the immediacy and eye mind assumption (i) the eyes tend to fixate those objects from which visual information is extracted to support their cognitive evaluation (e.g., Just & Carpenter, 1980; Rayner & Pollatsek, 1989). Additionally, (ii) fixation durations are agreed to reflect a reliable measure of how long processing of a particular stimulus takes (for a review see Rayner, 1998). Thereby, evaluation of participants' eye fixation behaviour may be a valuable tool to dissociate cognitive processes in numerical cognition.

However, in contrast to other neuro-cognitive domains such as e.g., reading research only few studies in the domain of numerical cognition have yet tapped the potential of eye-tracking data. Nevertheless, the eye-tracking methodology has already been employed to study basic as well as more complex numerical processes. First evidence for the validity of eye fixation behaviour for the evaluation of basic numerical processes has come from studies on basic number related tasks such as number reading (Brysbaert, 1995) and magnitude comparison (Moeller, Fischer, Nuerk, & Willmes, 2009a) as well as more complex tasks such as the number bisection task (Moeller, Fischer, Nuerk, & Willmes, 2009b) . As regards addition, a first study evaluating eye fixation behaviour has been reported by Green et al. (2007). Yet, the authors employed eye-tracking to validate the use of strategies, which the participants were instructed to apply, but did not use the eye fixation data to investigate the nature of the cognitive mechanisms underlying the difficulty of carry operations in multi-digit addition. This will be the main focus of the current study.

In summary, eye movement measures seem to be well-suited to offer new insights into the sequence and the nature of cognitive processes employed in both rather basic and more complex numerical tasks. Based on this the current study aimed at dissociating two different processes associated with solving carry addition problems. These processes may help to better understand what makes carry addition problems more difficult than their non-carry counterparts.

Objectives of the current study

Despite the fact that it is widely agreed that the requirement of a carry is a crucial predictor of difficulty in multi-digit addition, the question what exactly causes the difficulty associated with the carry operation is not yet resolved. We suggest that at least two processes involved in processing a carry in addition can be differentiated which may drive the increased difficulty of these problems.

(i) Before any carry procedure can be executed it has to be recognized that a carry is needed to compute the correct result. One way to determine whether a carry is needed or not is to keep track of the sum of the to-be-added unit digits: a carry is required whenever the unit sum is equal or larger than 10. In this case, the decade digit of the unit sum has to be carried to the decade digit of the result. Consider here for instance 25 + 39: the sum of the unit digits is 5 + 9 = 14 and thus a carry is necessary to obtain the correct result. Thereby, the necessity of a carry is driven by calculation processes upon the unit digits.

(ii) After it has been recognized that a carry is needed the carry procedure still has to be executed properly: This means that the carry, i.e., the decade digit of the unit sum, has to be added to the sum of the decade digits of the summands. In above mentioned example, the decade digit of the unit sum is 1 (from 14). Adding this to the sum of the decade digits of the summands, i.e., 2 + 3 + *1* = 6 (italics indicating carry) results in the decade digit of the correct result (25 + 39 = 64).

The carry effect in addition has mainly been investigated by reaction time experiments (Deschuyteneer et al., 2005; Fürst & Hitch, 2000; Klein et al., submitted) as well as functional Magnetic Resonance Imaging (fMRI) studies (Kong et al., 2005). However, to investigate the nature and sequence of the basic cognitive processes underlying the carry effect in addition a more fine-grained analysis by a supplemental evaluation of participants' eye fixation behaviour may be informative since eye fixations are an established indicator of what is being processed at the moment (e.g., Rayner & Pollatsek, 1989). Applied to an addition problem this means that identifying which digit is fixated at a time also indicates which digit is actually processed.

In the current study, we applied this assumption to an addition verification task and derived specific hypotheses for the influence of underlying cognitive processes on the eye fixation behaviour in addition problems with and without carry. First, the validity of eye movement data can be studied for this effect. Eye movement research has shown that increasing task difficulty is resolved by either more fixations and/or longer total reading times (see Rayner, 1998 for a review). Transferred to the addition task this implies longer total reading times in carry as compared to non-carry problems.

However, as already outlined above, the goal of the current study was not a mere replication of RT data. Rather, following up on above considerations on two possible origins of difficulty in carry addition problems, more specific hypotheses for the eye fixation behaviour can be derived:

(i) When the difficulty associated with a required carry operation arises at the unit calculation stage then longer total reading times should be observed on the unit digits of the summands for carry addition problems than for non-carry problems. Additionally, regressions from the second to the first summand should specifically involve the unit digits. This should be the case as specific evaluation of the units (i.e., the unit sum being equal or larger than 10) determines whether or not a carry is required.

(ii) When difficulty is associated with the execution of the carry procedure itself, then longer total reading times on the decade digit of the result should be found as the sum of the decade digits of the summands has to be updated by the carry to obtain the correct result (e.g., adding 1 to the sum of the decade digits). Accordingly, regressions from the result back to the summands should be prominently focused on the decade digits.

The present study pursued these questions.

METHOD

Participants: 20 students (15 female) of the Paris-Lodron University of Salzburg participated in the study as partial fulfilment of course requirements. Mean age was 22.7 years with a standard deviation (SD) of 2.4 years (range: 20 - 31 years). All participants reported normal or corrected to normal vision.

Apparatus: Eye fixation behaviour was recorded online using an EyeLink 1000 eye-tracking device (SR Research, Mississauga, Ontario, Canada) and stored for offline analysis. The EyeLink 1000 provides a spatial resolution of less than 0.5 degrees of visual angle at a sampling rate of 1000 Hz. Stimuli were presented on a 21'' monitor with resolution set to 1024 x 768 pixels and driven at a refresh rate of 120 Hz. Participants' heads were placed in a chin rest throughout the experiment to keep viewing distance (50 cm) and viewing angle constant.

Task, Stimuli, and Design: Participants had to evaluate whether displayed addition problems were solved correctly by simultaneously presented probes or not. The addition problems consisted of one- and two-digit numbers with a sum not exceeding 99. The stimulus set comprised 96 addition problems. In a factorial 2 x 2 design *problem size* (small vs. large, i.e. sum < 40 vs. sum > 40) and the requirement of a *carry operation* (required vs. not required) were manipulated orthogonally. Each problem was presented twice once together with the correct result and once with an incorrect probe. Incorrect probes deviated from the correct result by either 2 or 10 to prevent parity based solution strategies with average split kept constant (i.e., zero) across item groups (for an overview of stimulus properties see the Appendix). Addition problems were presented in the form xx + xx = xx at a position slightly to the right of the centre of the screen. To ensure that all participants started to encode the addition problem at the first summand a fixation point to the left of this summand was used (x/y coordinates: 112/384). Digits and arithmetic symbols were shown in white against a black background using the unproportional font *New Courier* (size: 48; style: bold). At this size each digit subtended 1.4 degrees of visual angle in width and 1.9 degrees in height. The distance between the both summands as well as between the second summand and the result equalled 10.4 degrees of visual angle to keep influences of parafoveal preprocessing at a minimum.

Procedure: After being seated with their head stabilized by a chinrest, participants were instructed to evaluate as fast and as accurate as possible whether the subsequently displayed addition problems were solved correctly or not: When a problem was presented together with its correct solution the right button of a response device had to be pressed by the index finger of the right hand whereas a button press by the left index finger on a left button indicated that a problem was presented with an incorrect solution. Then a nine-point calibration of the eye-tracking system was conducted to maximize spatial resolution for each participant. Thereafter, participants had to evaluate ten practice trials to familiarize with display layout and task requirements. The experiment was set up in four blocks of 48 items each and lasted about 45 minutes.

Analysis: All subsequent analyses exclusively incorporated data from items presented with a correct solution probe. This restriction was necessary as incorrect solution probes may be rejected following non-computational strategies (e.g., matching of intermediate unit sum). Additionally, evaluation processes of incorrect probes may be driven by factors such as split between incorrect probe and correct result which do not play a role in the evaluation of correct solution probes as they do not even exist for these (e.g., Klein et al., 2009; Klein et al., submitted). Furthermore, Menon et al. (2002) were able to show that responses to correct and incorrect probes differ even in their neural correlates: In incorrect addition problems specific cortex areas were additionally recruited in left dorsolateral and ventrolateral prefrontal cortex – areas usually not associated with the processing of domain specific numerical information. Instead, these areas seem to play a role in monitoring contextual information (e.g., Cabeza, Locantore, & Anderson, 2003; Ranganath, Johnson, & D'Esposito, 2000) and the generation of alternative solutions to a given problem (e.g., Donohue et al., 2005; Goel & Vartarian, 2005). As we were specifically interested in processes necessary for deriving the correct solution to an addition problem we restricted analyses to those items involving a correct solution probe[1].

Classification performance in terms of response latencies and error rates was evaluated by a two-way ANOVA (analysis of variance) incorporating the factors *problem size* (small vs. large) and *carry* (required vs. not required). To approximate normal distribution error rates were arcsine transformed prior to the analysis.

For the analysis of the eye fixation behaviour only problems with a large problem size were recruited. In the group of problems with a small problem size there were several problems involving a single-digit summand (e.g. 9 + 14 = 23). As in these trials the number of decade digits may vary, focussing on the problems with a large problem size ensured that the subsequent evaluation of the distribution of fixations over tens and units was not confounded by the fact that sometimes there were no decade digits to look at. For the subsequent evaluation of eye fixation behaviour in terms of total reading times on either the decade or unit digit of the first summand, the second summand, and the result areas of interest were defined: Each digit was centred in an area of interest 75 pixels wide and 180 pixels high. Whenever a fixation fell within one of these interest areas it was considered as a fixation on the corresponding decade or unit digit. Subsequently, the total reading time for each interest area was computed and submitted to a 2 × 3 × 2 ANOVA (analysis of variance) with the factors *carry* (required vs. not required), *problem element* (first summand vs. second summand vs. result) and *digit identity* (decade vs. unit digit). Whenever necessary, post-hoc comparisons were conducted using the Games-Howell test at a significance level of $p < .05$ to account for differing variances. Additionally, in the cases the sphericity assumption of the ANOVA was violated the Greenhouse-Geisser coefficient (GG) is given to allow for an

[1] Please note that we are well aware that analyzing the data of the incorrectly solved problems may be informative on possible strategy differences between processing these and correctly solved problems and thus, may further substantiate the observations by Menon et al. (2002). However, for the sake of brevity and lucidity the current study focused on evaluating processes underlying the carry effect in correctly solved addition problems.

adjustment of the degrees of freedom. To further investigate in what way a carry operation influenced the number of regressions (i.e., leftward saccades to a previous interest area) initiated from either tens or units of both the second summand as well as the result a 2 x 2 ANOVA was conducted discerning the factors digit identity (decade vs. unit digit) and carry (required vs. not required). Finally, despite the origin of the regressions also the target of the regressions was appraised. For regressions from the second summand to the first summand a 2 x 2 ANOVA with the factors digit identity (decade vs. unit digit) and carry (required vs. not required) was conducted while regressions originating from the result were analyzed by a 2 x 2 x 2 ANOVA involving the factors problem element (first vs. second summand), digit identity (decade vs. unit digit), and carry (required vs. not required).

RESULTS

Participants who performed at or below chance level in at least one of the experimental conditions where excluded from further analyses. This affected one participant. For the other participants error rates ranged from 1.0 % to 13.3 % with the mean at 7.0 % and a standard deviation of 3.4 %. Only response latencies followed by a correct classification were considered for the RT analysis. Additionally, a trimming procedure eliminated all latencies shorter than 200 ms and longer than 5000 ms in a first step. Subsequently, all RTs falling below or above three standard deviations of an individual participant's mean were excluded from the analyses in a second step. For the analyses of the eye fixation data only data from trials already included in the RT analysis were considered.

Classification performance

<u>ANOVA:</u> The ANOVA showed that both response latencies as well as error rates were reliably influenced by problem size [RT: $F(1, 18) = 86.26$, $p < .001$; errors: $F(1, 18) = 16.09$, $p < .001$, see Figure 1, Panel A]: addition problems with a relatively larger problem size

resulted in longer latencies (2772 ms) and a higher error rate (9.6 % errors) as compared to problems with a small problem size (1730 ms and 4.3 % errors). Moreover, carry also had a consistent influence on RTs and error rates [RT: $F(1, 18) = 121.92$, $p < .001$; errors: $F(1, 18) = 5.13$, $p < .05$, see Figure 1; Panel B] with increased latencies and more errors being associated with addition problems which require a carry operation (2471 ms and 9.3 % errors) than with problems not requiring a carry (2031 ms and 4.7 % errors). Finally, a significant interaction of problem size and carry was present for latencies [$F(1, 18) = 8.35$, $p < .01$] but not for error rates [$F(1, 18) < 1$], indicating that the carry effect for RTs was more pronounced for addition problems with a large problem (540 ms) size than for problems with a small problem size (341 ms).

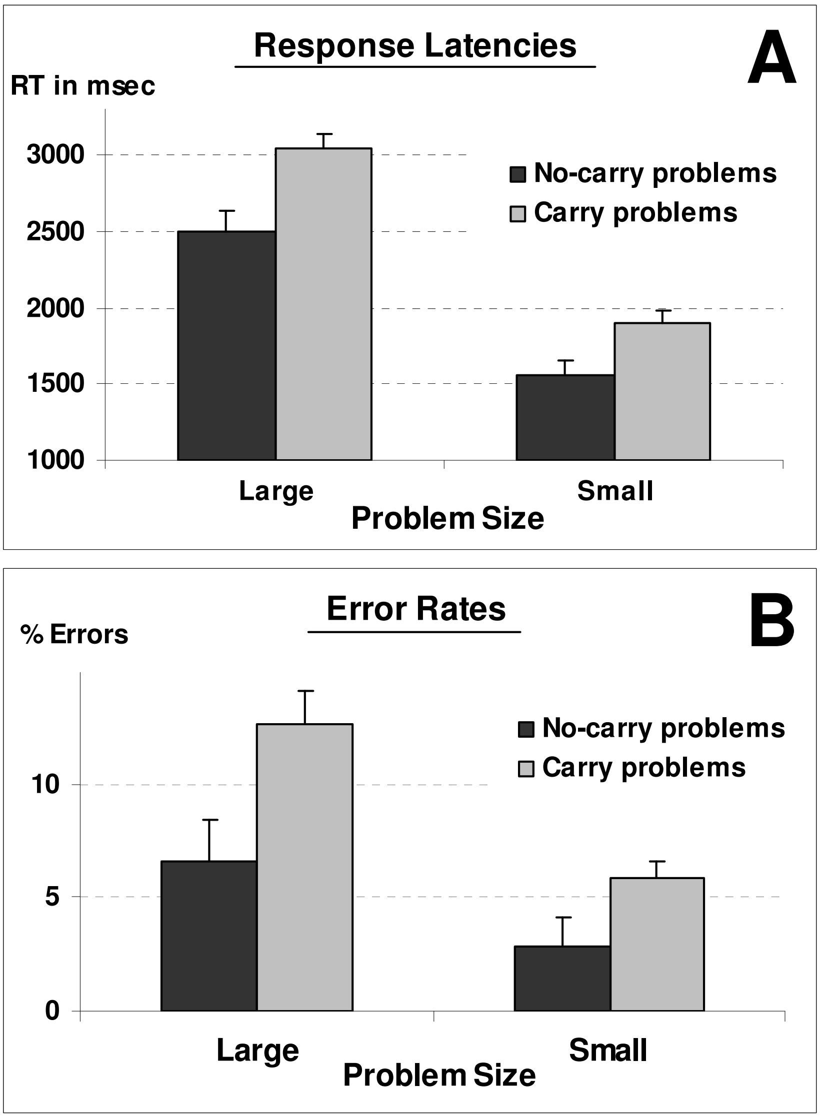
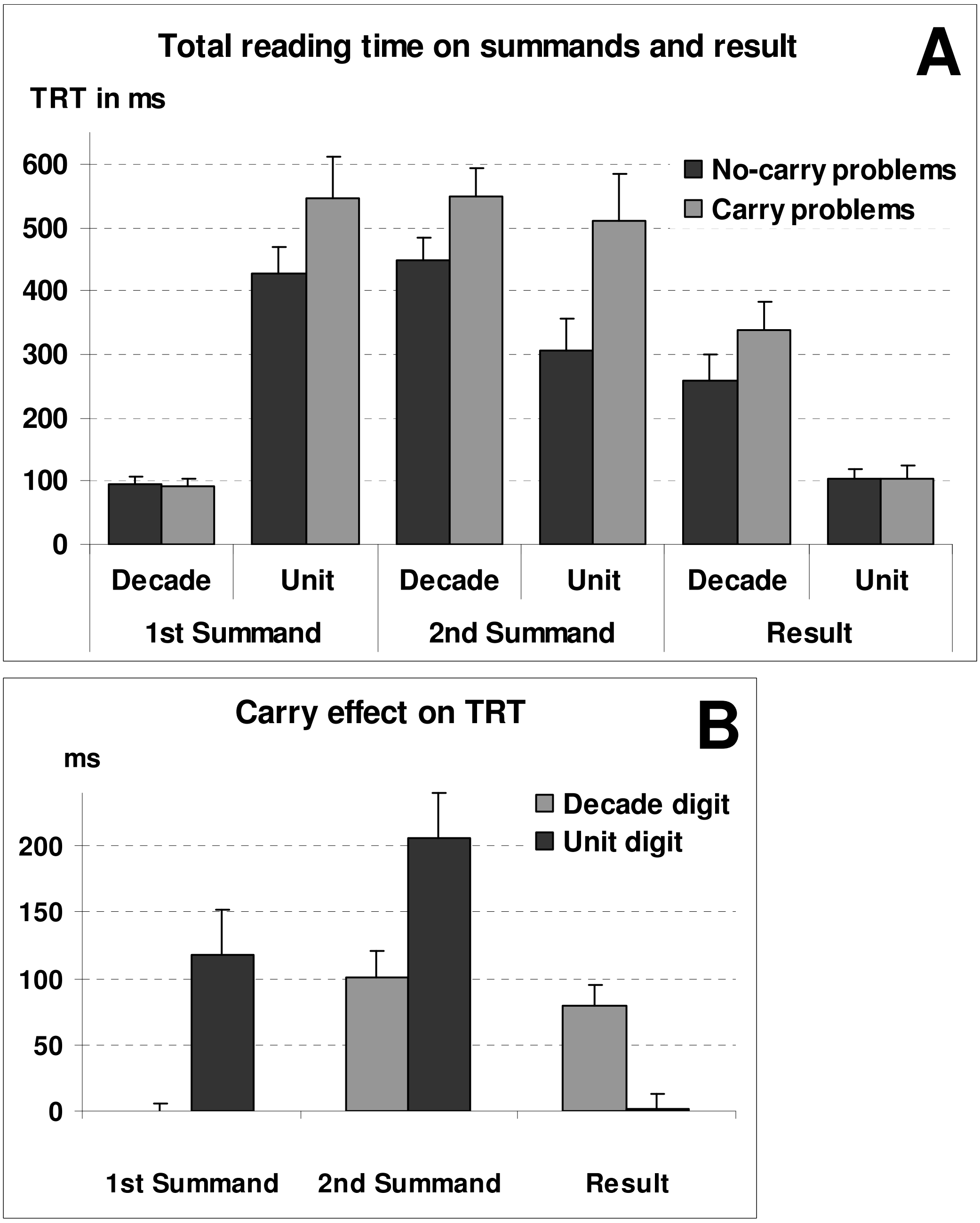


Figure 1: *Response latencies (Panel A) and error rates (Panel B) separated for carry and non-carry problems with either large or small problem size. Error bars depict 1 Standard Error of the Mean (SEM).*

Eye fixation behaviour



<u>Figure 2:</u> Distribution of total reading times across decade and unit digits of the summands and the result separated for carry and non-carry problems (Panel A). The carry effect in ms total reading time (i.e., TRT carry - TRT non-carry) on either of the summands or the result is shown in Panel B. Please note that the carry effect on the decade digit of the first summand as well as the unit digit of the result is actually not zero (i.e., - 1 ms and + 2 ms, respectively), but too small to be depicted in the figure at this resolution. Error bars reflect 1 SEM.

Total reading time on target[2]*:* The ANOVA discerning the factors carry (required vs. not required), problem element (first summand vs. second summand vs. result), and digit identity (decade vs. unit digit) revealed reliable main effects of carry [$F(1, 18) = 42.29$, $p < .001$] and problem element [$F(2, 36) = 35.47$, $p < .001$, see Figure 2, Panel A]. This indicated that addition problems requiring a carry operation were associated with a higher average total reading time per digit as compared to non-carry problems (354 ms vs. 267 ms, respectively). Additionally, fixations were not distributed equally between the two summands and the result. Post-hoc evaluations by the Games-Howell test indicated that total reading times were longest on the second summand followed by the first summand and the result (450 ms vs. 293 ms vs. 187 ms, respectively). The main effect of digit identity was not significant [$F(1, 18) = 1.40$, $p = .25$]. Moreover, the interaction of carry and problem element was reliable [$F(2, 36) = 23.33$, $p < .001$, see Figure 2, Panel A]. Post-hoc comparisons by the Games-Howell test showed that the increase in total reading time due to a required carry operation was most pronounced on the second summand but did not differ between the first summand and the result (+ 156 ms vs. + 64 ms vs. + 44 ms, respectively). Furthermore, problem element and digit identity interacted significantly [$F(2, 36) = 26.36$, GG = 0.81, $p < .001$]. Post-hoc testing by the Games Howell test indicated that the effect of digit identity (TRT on the unit digits – TRT on the decade digits) was largest on the first summand but did not differ between the second summand and the result (+ 399 ms vs. – 114 ms vs. – 170 ms, respectively). Additional *t*-tests revealed that upon the first summand the unit digits were fixated reliably longer than the decade digits (492 ms vs. 93 ms, respectively; $t(18) = 7.22$, $p < .001$), whereas this pattern was reversed for the result. Here, longer TRT were observed on the decade digit than on the unit digit (272 ms vs. 102 ms fixations, respectively; $t(18) = 4.35$, $p < .001$). However, for the second summand no significant difference between the total reading times on either the unit

[2] Please note that the analyses of the eye fixation behaviour were based on addition problems with a large problem size only. As addition problems with a small problem size also involved single-digit numbers the distribution of fixations across decade and unit digits would have been biased for these trials.

or the decade digits was found (393 ms vs. 507 ms, respectively; $t(18) = 1.46$, $p = .16$). Finally, the two-way interaction of carry and digit identity was significant [$F(1, 18) = 4.85$, $p < .05$] indicating that the requirement of a carry led to a reliably stronger increase of total reading times on the unit digits than on the decade digits of the problem (+ 111 ms vs. + 65 ms). Interestingly, this two-way interaction was further specified by the reliable three-way interaction of carry, problem element and digit identity [$F(2, 36) = 12.40$, $p < .001$]. Breaking down this interaction in its constituting two-way interactions revealed a significant interaction of carry and digit identity for the first summand [$F(1, 18) = 13.58$, $p < .01$] and the result [$F(1, 18) = 24.35$, $p < .001$] while it was only marginally significant for the second summand [$F(1, 18) = 4.46$, $p = .06$]. However, a closer inspection of the marginal means showed that for the first summand as well as for the second summand a required carry operation resulted in a more pronounced increase of total reading time on the unit digits compared to the decade digits (first summand: + 119 ms vs. ± 0 ms; second summand: + 205 ms vs. + 101 ms, see Figure 2, Panel B). Contrarily, for the result the carry effect on total reading time was stronger on the decade digit than on the unit digit (+ 80 ms vs. + 2 ms fixations, respectively, see Figure 2, Panel B).

Origin of regressions: The ANOVA evaluating the number of regressions from the second summand back to the first summand showed that carry addition problems were associated with a reliably higher number of regressions [0.87 vs. 0.61 regressions; $F(1, 18) = 27.58$, $p < .001$]. Moreover, the significant interaction of carry and digit identity [$F(1, 18) = 4.26$, $p = .05$] indicated that, in particular, the number of regressions from the unit digit of the second summand back to the first summand (as compared to regressions from the decade digit of the second summand) was increased in carry addition problems (+ 0.37 vs. + 0.15 regressions). The main effect of digit identity was not significant [$F(1, 18) = 2.37$, $p = .14$].

Comparably, also the number of regressions from the result back to the summands was moderated by the requirement of a carry operation [$F(1, 18) = 8.68$, $p < .01$]: in carry addition problems participants looked back to the summands significantly more often than in non-carry problems (0.54 vs. 0.46 regressions, respectively). Additionally, the main effect of digit identity indicated that more regressions to the summands started from the decade digit as compared to the unit digit of the result. This was further specified by the reliable two-way interaction of carry and digit identity [$F(1, 18) = 17.24$, $p < .001$]. The requirement of a carry specifically increased the number of regressions from the decade digit of the result back to the summands (+ 0.15 vs. ± 0 regressions).

Target of regressions: Examining the regressions from the second summand back to the first summand, the ANOVA showed a reliable effect of carry [$F(1, 18) = 11.72$, $p < .01$] with more regressions to the first summand for carry addition problems as compared to non-carry problems (0.58 vs. 0.44 regressions, respectively). Additionally, the main effect of digit identity was significant as well [$F(1, 18) = 42.20$, $p < .001$] indicating that the majority of regressions was directed to the unit digit when looking back to the first summand (0.83 vs. 0.19 regressions). Furthermore, the reliable interaction of these two factors [$F(1, 18) = 6.55$, $p < .05$] implied that the increase of regressions towards the first summand due to the requirement of a carry was more pronounced on the unit digit than on the decade digit of the first summand (+ 0.13 vs. + 0.05 regressions, respectively).

In line with these results, the ANOVA evaluating the regressions from the result back to the two summands also revealed a significant carry effect [$F(1, 18) = 27.22$, $p < .001$]: more regressions from the results to the summands were observed for carry than for non-carry problems. Furthermore, also the two main effects of problem element [$F(1, 18) = 18.42$, $p < .001$] and digit identity [$F(1, 18) = 26.21$, $p < .001$] were reliable. This indicated that more regressions were directed to the second summand as compared to the first summand (0.37 vs. 0.10 regressions, respectively) as well as to the decade than to the unit digits (0.31 vs. 0.16 regressions, respectively) of both summands. Moreover, all possible two-way interactions were statistically reliable but not the three-way interaction [$F(1, 18) < 1$]. The interaction of problem element and carry [$F(1, 18) = 21.27$, $p < .001$] indicated that the increase of regressions to the summands due to a carry was larger for the second summand than for the first summand (+ 0.17 vs. ± 0.00 regressions). Additionally, the interaction of summand and digit identity [$F(1, 18) = 12.71$, $p < .01$] meant that the effect of digit identity (i.e., more regressions to the decade digit) was more pronounced on the second summand (+ 0.26 vs. + 0.03 regressions). Finally, the interaction of carry and digit identity [$F(1, 18) = 4.50$, $p < .05$] implied that the increase of regressions towards the two summands due to a required carry was stronger for regressions towards the decade as compared to the unit digit (+ 0.12 vs. + 0.05 regressions, respectively).

In summary, the present results indicated dissociable influences of the requirement of a carry operation on the eye fixation patterns on the summands and the result in the addition verification paradigm employed. On the one hand, the need of a carry operation specifically increased total reading times on the *unit* digits of the summands (in particular on the unit digit of the second summand), possibly reflecting processes related to evaluating whether of not a carry is required. On the other hand, longer total reading times were observed on the *decade* digit of the results in carry addition problems assumed to indicate processes of carry execution. These results were further substantiated by the results of the analyses of origin and target of the observed regressions.

DISCUSSION

The current study set off to investigate the underlying basic processes from which the carry effect arises. On the one hand, specifically increased total reading times on the unit digits of the summands were hypothesized for carry as compared to non-carry problems. Moreover, regression from the second back to the first summand should specifically involve the unit digits as well. These findings would suggest an association of the carry effect with recognizing that a carry operation is needed which would be realized by specific evaluation of / calculations upon the units to derive whether or not a carry is required (i.e., is the unit sum being equal or larger than 10?). On the other hand, the difficulty of carry addition problems may be associated with the execution of the carry procedure itself. In this case we hypothesized to observe increased total reading times on the decade digit of the result for carry compared to non-carry problems as the sum of the decade digits of the summands has to be updated by the carry to obtain the correct result (e.g., by adding 1 to the sum of the sum of the decade digits). Based on these considerations, regression from the result back to the summands should also primarily involve the decade digits. The present data were meaningful on both aspects suggesting that the carry effect in multi-digit addition seems to originate from both increased processing demands for unit based calculations (to determine whether a carry is needed) as well as to finally execute the carry procedure rather than being a question of either/or. Results on each of these two issues and their implications will be discussed in turn.

Unit based calculations indexing that a carry is required

In line with our hypothesis that the recognition of a carry may be a factor determining problem difficulty we observed that the requirement of a carry specifically increased total reading times on the unit digits of the summands. This finding seems reasonable as it is the sum of exactly these unit digits which indicates whether a carry is needed, i.e., for all unit sums equal or larger than 10. Thus, the current data suggest that while the problem is encoded the requirement of a carry is determined by unit-based calculations as soon as it becomes apparent that the sum of the unit digits becomes a two-digit number. This interpretation is corroborated by a number of further observations. First, when assuming that participants performed the addition task from left to right (as implicated by the fixation sign on the left of the first summand) the necessity of a carry can be recognized first when encoding the unit digit of the second summand. Therefore, the increase of total reading times on the unit digits of the summands should be more pronounced on the unit digit of the second summand than on the unit digit of the first summand because no inference on a possibly required carry can be drawn from encoding the unit digit of the first summand only. The eye fixation pattern confirmed this hypothesis. The increase of total reading times on the unit digit of the second summand was indeed more pronounced than the corresponding carry effect on the unit digit of the first summand. Finally, additional evaluation of the number of regressions, in particular from the second back to the first summand substantiated this interpretation. Not only that the requirement of a carry led to a specific increase of regressions from the unit digit of the second back to the first summand, it was also observed that upon the first summand the majority of regressions was directed to the unit digits and particularly so when a carry was required. This again corroborates the notion that (i) unit-based calculation processes may be applied to evaluate whether a carry is required or not and (ii) the requirement of a carry specifically increases processing demands as regards the unit digits of the summands. As the former can only be decided while processing the unit digit of the second summand (before any regressions occurred), this suggests that gaze duration on the unit digit of the second summand (i.e., the sum of all fixation durations from first entering a region to first leaving it) should be sensitive to the size of the unit sum (see Brysbeart, 1995 for the sensitivity of gaze duration to magnitude processing). A correlation analysis confirmed this assumption. As the sum of the unit digits increased, gaze duration on the unit digit increased as well on the second summand but not on the first summand [$r = .68$, $p < .001$, $n = 48$; $r = - .22$, $p = .13$, $n = 48$, respectively].

Executing the carry – rule-based updating of the decade digit of the result

Moeller and colleagues (2009b) proposed a two-stage processing model of eye fixation behaviour for numerical tasks. In an initial bottom-up processing stage (indicated by influences on gaze duration) the constituting digits of the first number may be identified largely automatic and assigned their stimulus-driven lexical values (e.g., place-value position) by extracting their physical features and activating the memorized visual number forms (cf. Henik & Tzelgov, 1982; Moeller et al., 2009b). After integrating the constituting digits of a number into the place-value structure of the Arabic number system, the magnitude of the number may be accessed. The second processing stage reflecting a subsequent top-down driven wrap-up stage is supposed to be initiated after all the numbers of the actual problem have been encoded. At this stage, the numbers of an arithmetical problem may be integrated and put into relation with the other numbers by checking top-down mediated plausibility and/or processing rules using the lexical attributes of the single numbers. This process is supposed to be reflected by eye-movement measures capturing rather late processing such as total reading time. Based on their model of eye fixation behaviour in numerical tasks Moeller and colleagues (2009b) predicted longer total reading times on the decade digit of the result in carry addition problems than in non-carry problems. The authors suggested that calculating the sum of the decade digits and, in particular, updating this sum by the carry in carry addition problems (e.g., 25 + 39 = 64, units: 5 + 9 = 14; tens: 2 + 3 + 1 = 6) represents the application/execution of a specific processing rule. This should be reflected in eye movement measures associated with rather late top-down processing such as the total reading time on a given interest area.

In line with this prediction, total reading times on the decade digit of the result were specifically increased in carry as compared to non-carry addition problems. This finding indeed corroborated the interpretation that the execution of the carry procedure may represent the application of a procedural rule (i.e., "when the sum of the unit digits is equal or larger than 10 add the decade digit of the unit sum to the sum of the decade digits of the summands"). As for the unit-based calculations upon the unit digits of the summands, inspection of the regressions was also informative on processes related to executing the carry. In line with the argument that the execution of the carry may be specifically associated with processing the decade digit of the result (seeming to reflect processes of updating the decade sum of the summands by the carry) it was observed that most regressions from the result back to the summands started from the decade digit. Moreover, it was found that most of these regressions were directed to the decade digits of the summands and to the decade digit of the second summand, in particular. On the one hand, this corroborates the interpretation that processes concerning the execution of a carry in an addition verification paradigm involved specific processing of the decade digit of the result and its constituents, i.e., the decade digits of the summands. On the other hand, the results of analyzing the regressions also indicate that the increase of total reading time as observed upon the decade digit of the second summand may be driven by reinspections of this particular digit to come to the correct decade digit of the result in carry addition problems (see Figure 2 B) rather than being associated with initially evaluating whether a carry is needed or not.

Taken together, the present results are meaningful in three respects: First, in line with the results of Moeller et al. (2009a) the unbalanced distribution of fixations between tens and units corroborated the notion of tens and units being processed separately rather than integrated as the holistic representation of the whole number they constitute. Second, the results of analyzing the regressions indicated that, in the current paradigm, processes associated with actually calculating the correct result may not be finished completely before participants processed the solution probe and decided whether it was the correct solution to the problem at hand or not. Finally, and most importantly for the research question at hand, the data suggested that it does not seem to be an either/or distinction between unit based calculations determining that a carry is required and its processing. Instead, the present observations on the increased processing demands associated with a carry operation indicated that the carry effect in multi-digit addition seems to originate from both specific processing of the unit digit of the summands to evaluate by calculation that a carry is needed as well as particular processes reflecting the execution of the carry procedure (mostly associated with the decade digit of the result).

CONCLUSIONS / PERSPECTIVES

Previous studies indicated that processes of unit-decade integration as required by a carry operation pose particular difficulty on mental addition. In particular, response latencies as well as error rates are increased for addition problems requiring a carry operation (e.g., 18 + 27) compared to problems not requiring a carry (e.g., 13 + 32, e.g., Deschuyteneer et al., 2005; Kong et al., 2005). However, the cognitive underpinnings of this carry effect remained unclear. The current study aimed at pursuing this question by tracking participants' eye-movements during addition problem verification. Analyses of the eye-fixation data revealed a differential influence of decade and unit digits on mental addition: On the one hand, the requirement of a carry operation particularly increased total reading times on the unit digits of the summands, whereas total reading times were particularly increased on the decade digit of the result on the other hand. Furthermore, this dissociation was accompanied by a specific increase of regressions from the units of the second to the units of the first summand as well as from the tens of the result to the tens of the summands in carry addition problems compared to non-carry problems. This implies that processing a carry is not a unitary process.

Instead, it seems to involve different aspects that can be differentiated by the evaluation of participants' eye fixation behaviour. The current data indexed that both correctly recognizing the requirement of a carry (i.e., evaluating whether the unit sum is equal or larger than 10) as well as its execution (i.e., updating the decade sum by the carry) seem to be associated with increased processing demands of carry addition problems as compared to non-carry problems. In the long run, there is the possibility that the finding of specific differences in the fixation patterns for carry and non-carry problems may be informative for understanding math impairments. For instance, evaluating the fixation pattern of children with poor mathematical abilities may offer the possibility to distinguish whether a required carry is just not recognized or rather not executed. Finally, the current study indicates that evaluating eye fixation behaviour in mental arithmetic can be a promising tool to identify the basic cognitive processes underlying complex behavioural effects such as the carry effect.

APPENDIX A

Stimulus properties of the used item set

	Small problem size		Large problem size	
	Non-carry	Carry	Non-carry	Carry
Summand 1	9.92	10.38	41.29	38.46
Summand 2	10.42	10.29	41.58	43.21
Problem size categorical	1	1	2	2
Carry yes/no	0	1	0	1
Correct sum (= problem size)	20.33	20.67	82.88	81.67
log correct sum	1.28	1.29	1.92	1.91
Decade sum	1.25	0.75	7.58	6.83
Unit sum	7.83	13.17	7.04	13.33
Smaller summand left	0.50	0.50	0.50	0.50
Parity summand 1	1.67	1.29	1.46	1.46
Parity summand 2	1.33	1.46	1.42	1.46
Parity correct sum	1.25	1.33	1.38	1.33
5 in unit position of summand 1	0.17	0.17	0.17	0.17
5 in unit position of summand 2	0.17	0.17	0.17	0.17
5 in unit position of correct sum	0.08	0.08	0.08	0.08
5 in decade position of summand 1	0.00	0.00	0.04	0.04
5 in decade position of summand 2	0.00	0.00	0.04	0.04
5 in decade position of correct sum	0.00	0.00	0.00	0.00
Decade of summand 1	0.58	0.38	3.75	3.21
Decade of summand 2	0.67	0.38	3.83	3.63
Unit of summand 1	4.08	6.63	3.79	6.38
Unit of summand 2	3.75	6.54	3.25	6.96
Decade of correct sum	1.25	1.75	7.58	7.83
Unit of correct sum	7.83	3.17	7.04	3.33
Incorrect solution probe (distractor)	22.00	22.33	84.54	83.33
log distractor	1.31	1.33	1.92	1.92
Distance between sum/distractor	-1.67	-1.67	-1.67	-1.67
Absolute value distance sum/distractor	6.00	6.00	6.00	6.00
Decade crossing sum/distractor	0.50	0.50	0.50	0.50

Section 3

On the importance of place-value integration in arithmetical development

Study 3

Language effects on children's non-verbal number line estimations

ABSTRACT

The mental number line of children is usually assumed to be language-independent; however, this independency has not yet been studied. In this cross-cultural study we examined the influence of language properties on a non-verbal version of the number-line task in Italian- and German-speaking first grade children. The essential difference between the two languages concerns the inversion property of most German multi-digit numbers (e.g., 48 → "eight-and-forty"), whereas in Italian number-words no inversion is found. The analysis revealed two language-specific differences in the number line task: (1) generally, the estimates of Italian children were more accurate than those of Austrian children, even when controlling for general cognitive abilities. (2) Italian children performed particularly better when inversion errors led to large estimation errors. In conclusion, these findings indicate that the organization of children's mental number-line is indeed influenced by language properties even in non-verbal settings.

INTRODUCTION

The understanding and processing of numerical quantities constitutes one of the most far reaching steps in the development of numerical cognition. It is not only decisive for performance in basic numerical tasks such as estimation or number comparison; but also it is thought to serve as a building block for competences in higher levels of numerical cognition such as calculation (e.g., Holloway & Ansari, 2008; Laski & Siegler, 2006). Thus, given the importance of this basic numerical competency, understanding their developmental trajectories and their cognitive underpinnings seems crucial. However, knowledge on how the representation of symbolic number magnitude develops and on the factors which determine this development (e.g., language) is still rather patchy.

The mental number line and its development

It is widely agreed that number magnitude processing operates upon a spatially organized representation of numbers along a mental number line (MNL; e.g., Dehaene & Cohen, 1995; Restle, 1970; but see van Opstal, Gevers, de Moor, & Verguts, 2008, for a different view). Upon this MNL, relatively smaller numbers are associated with the left side, whereas numbers are represented further and further to the right with increasing magnitude (e.g., Dehaene, Bossini, & Gireaux, 1993; Zorzi, Priftis, & Umiltà, 2002; Zorzi, Priftis, Meneghello, Marenzi, & Umiltà, 2006; Priftis, Zorzi, Meneghello, Marenzi, & Umiltà, 2006; but see Shaki, Fisher, & Petrusic, 2009; Zebian, 2005 for evidence on a reversed orientation of the mental number line in cultures reading from right to left). Commonly, this spatial mapping of numbers is thought to develop around the age of 7 (Berch, Foley, Hill, & Ryan, 1999; van Galen, & Reitsma, 2008; but see Opfer, Thompson, & Furlong, in press; Opfer & Furlong, this issue for evidence suggesting first spatial-numerical associations even in preschoolers aged 4). However, the developmental trajectories of this spatial representation of number magnitude and the determining factors are still under debate (Booth, & Siegler, 2006; Ebersbach, Luwel, Frick, Onghena, & Verschaffel, 2008; Moeller, Pixner, Kaufmann, & Nuerk, 2009a; Barth & Paladino, in press).

The typical task to assess children's spatial representation of number magnitude requires them to determine the spatial position of a specific number on a physical line flanked by two numbers denoting the range of this hypothetical number line (e.g., 0-to-100; see Siegler & Opfer, 2003; Siegler & Booth, 2004; Booth & Siegler, 2006; Muldoon, Simms, Towse, Burns, & Yue, this issue; see also Opfer & Siegler, 2007; Opfer & Thompson, 2008 for the 0-to-1,000, and Thompson & Opfer, in press, for the 0-to-10,000 number range). In the context of this so-called number line task, children's ability to estimate a number's spatial position on this line is assumed to reflect their internal representation of numbers along the MNL. It is suggested that the coding of number magnitude in children changes gradually from logarithmic to linear representation as a function of age and experience (e.g., Booth & Siegler, 2008; Opfer & Siegler, 2007). This means that at first the spatial mapping of number magnitude is best conceptualized as being logarithmically compressed so that the perceived distances between two adjacent numbers on the MNL decrease as their magnitudes increase (e.g., Dehaene, 1992; 2001). With increasing age and experience the MNL is suggested to be organized in a linear fashion so that distances between two adjacent numbers become invariant to increasing magnitude (e.g., Gibbon & Church, 1981; Brannon, Wusthoff, Gallistel, & Gibbon, 2001).

However, two recent studies suggest the alternative view that performance in a number line task may not necessarily reflect a change from logarithmic to linear coding of number magnitude, but rather a change from a two-linear representation to a linear one (Ebersbach et al., 2008; Moeller et al., 2009a, see also Barth & Paladino, in press for an account based on a power model of proportion judgements). Using segmented regression analyses Ebersbach and colleagues (2009) observed that the breakpoint of the two-linear representation was correlated to children's counting abilities. The authors interpreted this finding to suggest two separate representations of numbers within and outside the children's counting range with an identifiable change point reflected by the breakpoint of a two-linear fitting function. Picking up on the idea of an initially two-linear representation of number magnitude, Moeller and coworkers (2009a) argued that performance changes with age and experience can also be interpreted as an improvement in integrating the single digits' magnitudes of tens and units in compliance with the place-value structure of the Arabic number system. That is, to correctly solve the number line task, one has to be aware that the distance between 0 and 40 is ten times as large as the distance between 0 and 4, as this is an attribute of the base-10 structure of the Arabic number system. However, when children have not yet acquired a comprehensive understanding of this relationship, they may consequently overestimate one-digit number intervals (i.e., misplacing them towards the right). Consequently, the remaining two-digit numbers (in the case of the 0-to-100 interval is assessed) have to be located in a compressed manner upon a relatively short segment of the hypothetical number line. This would suggest two linear representations in a number line task from 0-to-100 with a theoretically fixed breakpoint at 10: one for single digit numbers and one for two-digit numbers. In segmented regression analyses Moeller et al. (2009a) observed that a two-linear model with an assumed breakpoint at 10 (reflecting two separate representations for single- and two-digit numbers) fits the empirical data even better than a logarithmic model. Based on this intriguing finding, Moeller et al. (2009a) concluded that to estimate the magnitude of a given number in the number line task correctly, integration of the single digits' magnitude of tens and units complying with the place-value structure of the Arabic number system is inevitable. In this view the magnitude of a two-digit number is not represented holistically but built up by integrating tens and units into a coherent representation: This decomposed account states that the magnitude of the constituting digits is still represented even when only the overall magnitude of the number is task relevant; this has been demonstrated successfully in other tasks such as number magnitude comparison (see also Nuerk, Kaufmann, Zoppoth, & Willmes, 2004; Pixner, Moeller, Zuber, & Nuerk, 2009, for children data; Nuerk, Weger, & Willmes, 2001; 2004; Moeller, Nuerk, & Willmes, 2009b, for adult data; Nuerk & Willmes 2005, for a review) and has now also been applied to the number line task (cf. Moeller et al., 2009a). Taken together, Moeller et al. (2009a) specified the basic two-linear model as proposed by Ebersbach et al. (2008) by the assumption of a fixed breakpoint at 10 driven by theoretical consideration regarding the place-value structure of the Arabic number system as retained by decomposed processing of tens and units. Against this common background of the Ebersbach et al. (2008) and Moeller et al. (2009a) model one aim of the current study was to differentiate the notion of decomposed processing from the assumption of a holistic representation of overall magnitude as reflected by the logarithmic model.

However, the implementation of the place-value system differs in important aspects between the number word systems of various languages regarding for instance the order in which tens and units are referenced in number words (see Comrie, 2005, 2006 for a more detailed discussion). Thus, one might ask whether these specific differences may affect the acquisition of the mental representation of number magnitude. For example, more complex number word systems might be expected to impede this acquisition process whereas transparent number word systems, in particular regarding place-value structures, might ease the acquisition.

To further pursue this question, we will briefly review to what extent language specificities have been found to influence numerical understanding. Subsequently, it will be outlined in which way the notion of language-specific effects may affect results in a number line task.

Language specificities and their influence on number representation

In general, the idea of language-specific influences on numerical cognition is not new (Hunt & Agnoli, 1991; Miura, Okamoto, Kim, Steere, & Fayol 1993; Nuerk, Weger, & Willmes, 2005; Seron & Fayol, 1994; Zuber, Pixner, Moeller, & Nuerk, 2009; see also Krinzinger et al., this issue, for a comprehensive evaluation of language, sex and curricula on performance in a standardized mathematics achievement test for children from France, Belgium, Germany and Austria). More specifically, we aim to study the influence of inversion: In several number word systems (e.g., German, Dutch, Arabic, Maltese, Malagasy, etc.; Comrie, 2005), the order of tens and units in number words as compared to digital notation is inverted. Inversion means that tens and units are spoken in reversed order, for example 21 is spoken as "one-and-twenty". Contrasting calculation performance of French (non-inverted number words) and Dutch speaking adults (inverted number words) revealed that inversion seems to impede calculation performance in some conditions (Brysbaert, Fias, & Noël, 1998). Furthermore, inversion influences were even observed in magnitude comparison tasks in adults (Nuerk et al., 2005). Finally, inversion also affects number processing performance in children with the inversion property seeming to pose an additional obstacle on the acquisition of numerical abilities. For instance, transcoding performance of German speaking children has been found to be mainly influenced by language-specific attributes like inversion (Zuber et al., 2009).

However, the influences observed by Zuber and colleagues (2009) have been observed in a verbal task like transcoding: In their study, children had to write down Arabic numbers to dictation. Thus, there is a verbal component in this tasks and correct understanding of number words is absolutely necessary to perform such a transcoding task. In the case that a more complex number word structure would only influence tasks in which number words are directly involved, such a finding would not be very dramatic as number processing and/or calculation procedures with non-verbal symbols should not be affected by such a cultural specificity. Yet, in adults it has been observed that the inversion property influences even non-verbal tasks such as Arabic number comparison (Nuerk et al., 2005). In this non-verbal task, the integration of tens and units seemed to be more difficult in a language with inversion (i.e., German) than in a language without inversion (i.e., English). However, Moeller et al. (2009a) recently argued that the integration of the ten's and unit's magnitudes is also essential for the performance pattern in the number line task making the general objective of this study straightforward: When inversion influences unit-decade – integration into the place-value structure of the Arabic number system and when this integration is essential for performance in the number line task, inversion should also influence performance in the number line task. In particular, we should observe specific cross-cultural language differences in a task which was previously assumed to index (non-verbal) spatial magnitude representation of numbers.

Objectives

As outlined above, the current study aimed at investigating whether language as reflected by the structure of its number word system influences performance in a non-verbal spatial-numerical task. This issue was pursued by comparing German-speaking children using an inverted number word system to Italian-speaking children using a regular non-inverted number word system regarding their performance in a number line task. Thus, the inversion property of German number words is used to evaluate whether the reversed order of tens and units poses particular difficulty on children developing and/or applying the mental spatial representation of two-digit number magnitude.

More specifically, the hypotheses were three-fold:

- In a first step differences in overall estimation accuracy between Italian and Austrian (native German-speaking) children when performing the number-line task were of interest. It was expected that the reversed order of tens and units in German increases the estimation error in Austrian children.

- In a second step it was investigated whether apart from the log-to-linear assumption children's overall estimation performance may also be explained by a two-linear model proposing two separate representations of one- and two-digit numbers (cf. Moeller et al., 2009a). When there is evidence for the validity of such a two-linear model this would imply that developing a mental representation of number magnitude incorporates the mastery of the place-value structure of the Arabic number system.
- Finally, in a third step it was evaluated whether a possible confusion in assigning the correct value to the individual digits as determined by their position within the number (i.e., mixing up tens and units) can account for the general hypothesized poorer estimation performance of German-speaking children. In the context of place value integration German-speaking children's estimations should be particularly erroneous for numbers with a large inter-digit distance (e.g., 6, as in 82 which may be mixed up with 28, thereby resulting in a large estimation error) compared to numbers with a small inter-digit distance (e.g., 2, as in 54 which may be mixed up with 45, thereby leading to a rather small estimation error), as confusing tens and units is especially detrimental in these numbers. Related to this even a more specific hypothesis can be made. When the performance of German-speaking children confuse tens and units more often than Italian-speaking children possibly due to the inversion property of German number words, German-speaking children should specifically underestimate numbers such as 82 where mixing up tens and units results in a smaller number (→ 28). Contrarily, they should overestimate numbers such as 27 for which mixing up tens and units results in larger number (→ 72). As there is no inversion in Italian number words, this under-/overestimation should be more pronounced in German-speaking children.

METHOD

The reported experiment was part of a larger cross-cultural project investigating the early development of numeracy in children with different number words systems. Children in Italy and Austria were examined in a variety of numerical and non-numerical tasks. For the sake of brevity, this article will focus on the cross-cultural effects on performance in the number line task.

Participants: 130 German-speaking first graders (63 girls; see Moeller et al., 2009a) and 107 Italian speaking first graders (55 girls) participated in the study[1]. The German-speaking children were recruited from five Austrian elementary schools and were speaking German as their native language. Their mean age was 7 years 4 month (SD = 7.1 month). The Italian speaking children were recruited from two elementary schools in Italy, speaking Italian as their native language. Their mean age was 6 years 11 months (SD = 3.4 months). The age difference between the two samples was significant [$t(236) = 6.68$, $p > .01$]. The higher mean age of Austria children can be attributed to the fact that in Austria some children go to preschool first, thus entering first grade one year later. All children had normal or corrected to normal vision.

The study took place at the end of first grade. Please note that the curriculum regarding math education is virtually identical in Austria and Italy. In particular, children in both countries should have mastered the numbers up to 20 (including 0) under consideration of cardinality, ordinality as well as first arithmetic operations (i.e., addition and subtraction) within this range. Furthermore, basic numerical competencies such as quantity discrimination and matching as well as quantity comparisons should have been acquired.

[1] Please note that the German-speaking sample was identical to that investigated by Moeller et al. (2009a).

Children with an IQ (measured by CFT1; Catell, Weiss & Osterland, 1997) more than 1 SD below the average were excluded from the initial sample. This affected 2 children in Austria and 10 children in Italy.

Stimuli and design: The number-line task used was a paper-pencil version of a number-to-position task, requiring children to estimate the position of a given number on an empty number-line. Each stimulus of the task involved a 10-cm line with the left end labeled "0", and the right end labeled "100". The numbers to be estimated were printed centrally above the line in Arabic notation. Participants had to mark the position of the following numbers on the hypothetical number line in the given order: 27, 2, 64, 35, 7, 13, 99, 75, 47, 3, 11, 82, 95, 9, 17, 6, 18, and 53. Most important for the current study was the fact that no verbal number word was (explicitly) involved in the task neither at the presentation nor at the response level.

Procedure: Participants were tested in a one-to-one single session with the experimenter. Trials were presented sequentially one by one. Children were told to neither count nor use any other strategy but estimating the position of the presented number on the line. The number-line task started with an initial orienting problem, followed by the experimental problems. On the orienting problem, children were asked to estimate the position of the number 50 for practice purposes. Neither further information before the practice trial (e.g., indicating 50 to be the middle of the scale) nor feedback on performance after the practice trials or any of the critical trials was given.

Analysis: In a first step, the deviation of estimated position from true position for every single item was measured automatically to the nearest millimetre and the individual deviation error of the estimated from the actual value (in percent) was computed for each child separately to obtain estimation accuracy. To compare overall estimation accuracy between Italian and Austrian children an univariate ANCOVA with the factor language and the covariate *T*-value of CFT1 was computed.

In a second step, R^2 adjusted by the number of free parameters of the respective model (i.e., 2 in the case of the logarithmic model vs. 3 in the case of the two-linear model with fixed breakpoint) as a measure of goodness of fit independent of influences of model proximity (e.g., Kyllonen, Lohman, & Woltz, 1984) was computed for a linear and a logarithmic fitting function for each child individually (see also Appendix A for further discussion and analyses addressing the issue of possible model overfitting). Analogous to Moeller et al. (2009a), for the scale 0-100 adjusted R^2 for a segmented two-linear regression with the fixed break point at 10 was computed for each child individually. Mean adjusted R^2 values for the different fitting models (simple linear, logarithmic, and two-linear) for the Italian-speaking children were compared by paired samples *t*-tests. All pair wise comparisons were Bonferroni-Holm corrected (Holm, 1979) to account for alpha accumulation in multiple comparisons. Additionally, R^2 values were arcsine transformed prior to the analyses as they cannot be assumed to be normally distributed.

Finally, in a last step estimation performance for items with a small interdigit distance (i.e., 2; to-be-estimated numbers: 35, 53, 64, and 75) were contrasted to items with a large interdigit distance (i.e., 3-6; to-be-estimated numbers: 27, 47, 82, 95) to evaluate the origin of possible language differences. Please note that these two sets of items did not differ in problem size as reflected by the mean magnitude of the four numbers each [PS; $t(3) = 0.35$, $p = .75$; mean PS for small interdigit distance: 56.75; large interdigit distance: 62.75]. To substantiate the interpretation of the latter analysis it was evaluated whether the under-/overestimation possibly provoked when confusing tens and units of these numbers was more pronounced in German- as compared to Italian-speaking children. As both German- and Italian-speaking children underestimated above mentioned two-digit numbers (i.e., 27, 35, 47, 53, 64, 75, 82, and 95) we first computed the mean estimation error over all participating children and then subtracted each child's individual estimation error from it. A negative difference thus indicated that a child underestimated a given number even stronger than expected by the mean estimation error whereas a positive difference indicated that a child underestimated a given number less than expected by the mean estimation error, thus reflecting a relative overestimation. Additionally, to ensure that possible language differences were not driven by differences in the distribution of the computed difference scores between German- and Italian-speaking children a z-transformation of these individual difference scores was conducted. Subsequently, we directly tested the hypothesis that German-speaking children's under-/overshoot should be more pronounced and thus the difference between items provoking either of it should be reliably larger for them using a *t*-test.

RESULTS

CFT-1: The results of the CFT-1 are presented first as they have an influence on the following analyses. A *t*-test for independent samples revealed a significant difference of *T*-values in CFT-1 between the two countries, $t(223) = 10.32$, $p < .001$, with Italian children showing a significant lower average *T*-value (48.27, SD = 6.5) than Austrian children (58.59, SD = 8.1). We examined whether this effect was due to the fact that the Italian children were younger than the Austrian children. However, even when taking the age group instead of grade as reference in the CFT-1, Italian children still scored significantly lower. For that reason, individual *T*-values were included as a covariate in all further analyses.

Overall estimation accuracy:

An univariate ANCOVA on each child's mean percent absolute error revealed a reliable main effect of language for the 0-to-100 scale [$F(1, 223) = 7.97$, $p < .01$] with a mean estimation error of 17.78% for Italian and 21.06% for Austrian children (see Figure 1). This indicated that Italian children's estimates are in general more accurate than Austrian children's estimates. Additionally, the covariate (i.e., individual CFT-1 *T*-values) reached significance [$F(1, 223) = 13.83$, $p < .001$].

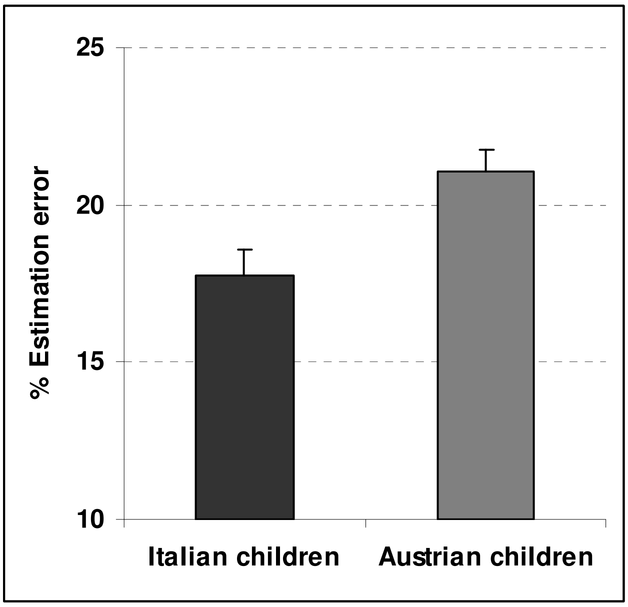


Figure 1: Mean estimation error in percent for Italian and Austrian children. Error bars reflect 1 Standard Error of the Mean (SEM).

Fit of simple linear, logarithmic and two-linear regression models:

The results of the Austrian children have already been described in great detail in Moeller et al. (2009). Therefore, the results relevant for the present article will only be summarized briefly. For the 0-to-100 scale logarithmic fitting was better as compared to the simple linear one [adj. $R^2_{\text{log}} = .70$ vs. adj. $R^2_{\text{lin}} = .61$]. However, when contrasting the logarithmic fitting to that of the two-linear model, the two-linear model accounted for a reliably larger part of the variance than the logarithmic model [adj. $R^2_{\text{two-lin}} = .81$ vs. adj. $R^2_{\text{log}} = .70$, see Figure 2, please note that for illustrating purposes the results of the item analysis are depicted]. Furthermore, inspection of the children's individual breakpoints showed that 95 out of 128 children assessed exhibited a breakpoint around 10 with the median of the breakpoints at 11.33. While this is a little bit above the proposed breakpoint at 10 it nevertheless corresponds nicely to the first two-digit numbers children had to position upon the hypothetical number line.

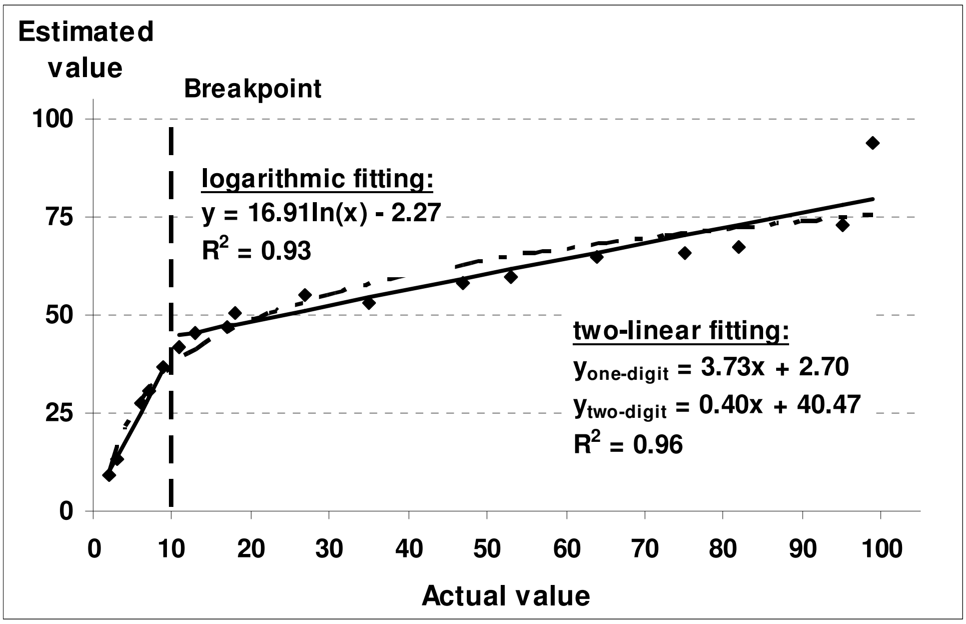


Figure 2: *Results of both logarithmic (dashed line) and two-linear (solid line) fitting for the mean estimates across all German-speaking children (see Moeller et al., 2009 for further details)*

Similar to the Austrian children, for Italian children logarithmic fitting accounted for a reliably larger part of overall variance than did simple linear fitting for the 0-to-100 scale [$t(96) = 3.46$, p < .01] with $R^2_{log} = .68$ and $R^2_{lin} = .63$. However, comparable to the results for the German-speaking children two-linear fitting was significantly better than the logarithmic fitting in terms of descriptive adequacy for Italian children as well [$t(96) = 11.73$; p < .001]. Average adjusted $R^2_{two\text{-}lin}$ was .82 compared to an R^2_{lin} of .68.[2]

Two more observations corroborate the superior fit of the two-linear model. First, as found for the German-speaking sample, inspection of the correlation between the slopes of the two linear segments of the two-linear model showed that these were negatively correlated (r = - .42; n = 97; see Moeller et al., 2009a for detailed explanations). Second, inspection of the distribution of the individual breakpoints of the participating children revealed another

[2] Please note that results were identical when excluding all teen numbers from the analyses which are constructed differently in German and Italian number words [Italian-speaking: $t(96) = 13.82$, $p < .001$, mean of individual R^2 for logarithmic fitting $R^2_{log} = .69$, mean $R^2_{two\text{-}lin} = .84$; German-speaking: $t(127) = 5.82$, $p < .001$, mean $R^2_{log} = .70$, mean $R^2_{two\text{-}lin} = .83$]. This suggested that the current results were not driven by differences at the level of teen numbers which are a special case in both languages.

striking similarity between the original German-speaking participants and the Italian children. Again, the vast majority of children (71 out of 97 Italian-speaking children) had a breakpoint around the hypothesized change from one- to two-digit number representation at 10 (see Figure 3). This is also reflected by the median of the individual breakpoints equaling 9.82 (M = 21.19, SD = 45.95).

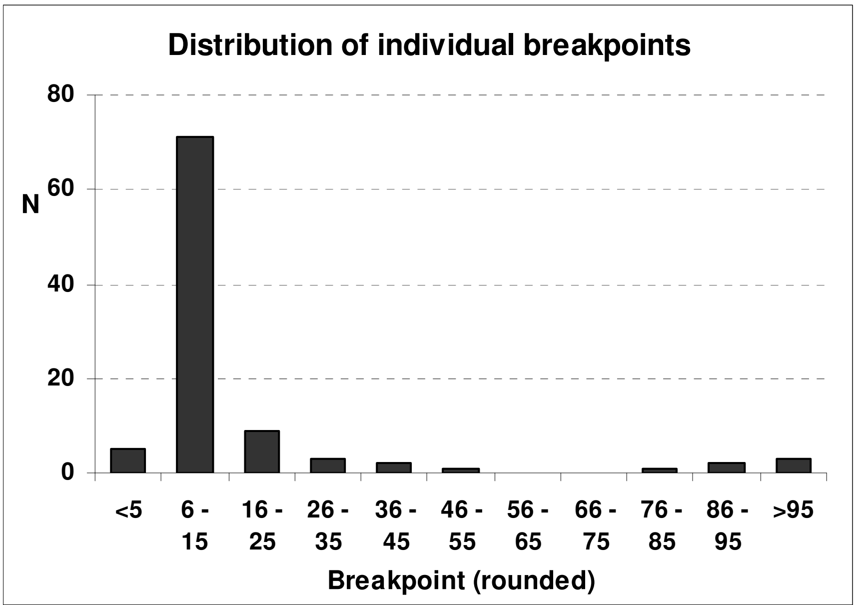


Figure 3: *Distribution of individually computed breakpoints (rounded to the nearest integer) and arranged in clusters around the multiples of 10.*

Taken together, our results strengthen the assumption of two separate but linear representations for singe-digit and two-digit numbers as both Austrian and Italian children's estimates on the 0 - 100 scale are accounted for best by a two-linear model with an assumed break point at 10 (see Figure 4 for an illustration of regression models based on the mean estimation of all children). This also implies that the representations of two-digit numbers may be an integration of the two single digits constituting the number. As German and Italian differ in the way two-digit numbers are verbalized (see above) we were interested whether the performance differences observed for the 0 - 100 scale were specifically driven by the inversion properties of the German number-word system, albeit an identical underlying representation of two-digit numbers for German-speaking and Italian-speaking children as assessed in the number line task

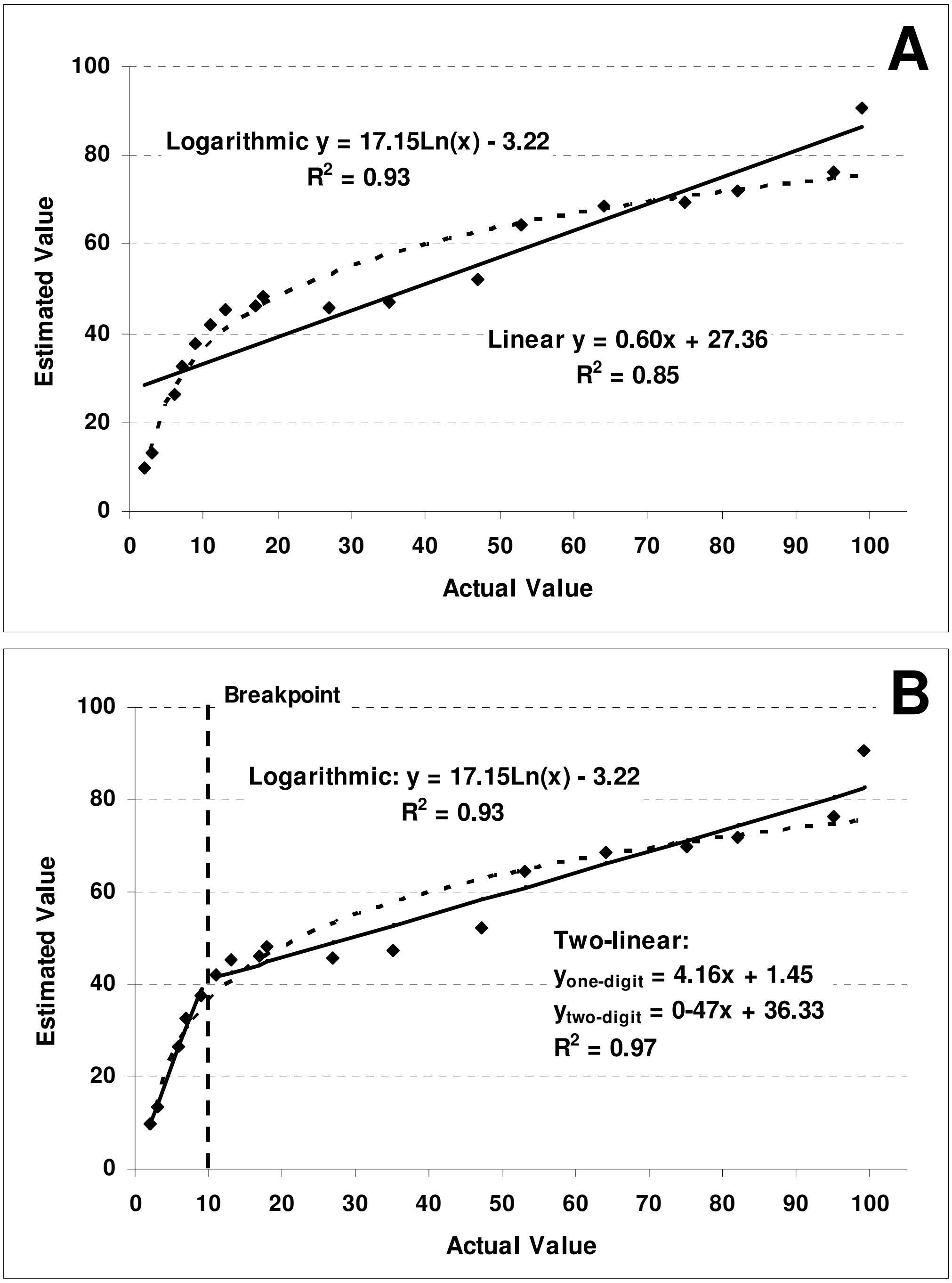


Figure 4: *Regression lines based on the mean estimates of all children per data point showing the differences of overall logarithmic (dashed line) versus simple linear (solid line, Panel A) and two-linear fitting (solid line, Panel B) of the 0-to-100 scale. Please note that the data presented here reflects an item analysis and thus does not directly correspond to the participant-based analyses previously described. These data were chosen to be given as they allow for a better illustration of differences and similarities of the logarithmic and the two-linear model. Please also note that an additional two-linear regression in which the optimal break-point was computed instead of using the theoretically driven breakpoint at 10 revealed an identical adjusted R^2 of .97 with the optimal break-point at 9.57.*

Inversion influences on estimation accuracy

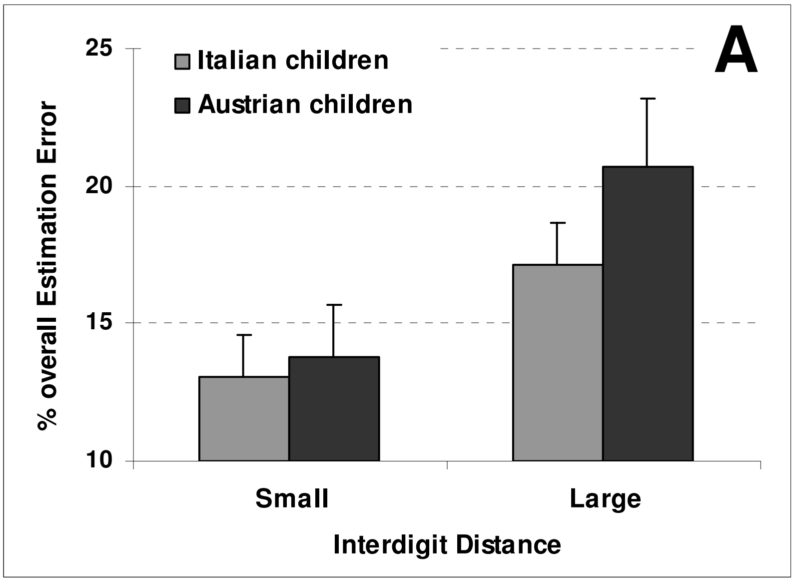
Estimation performance for items with large and small inter-digit distance:

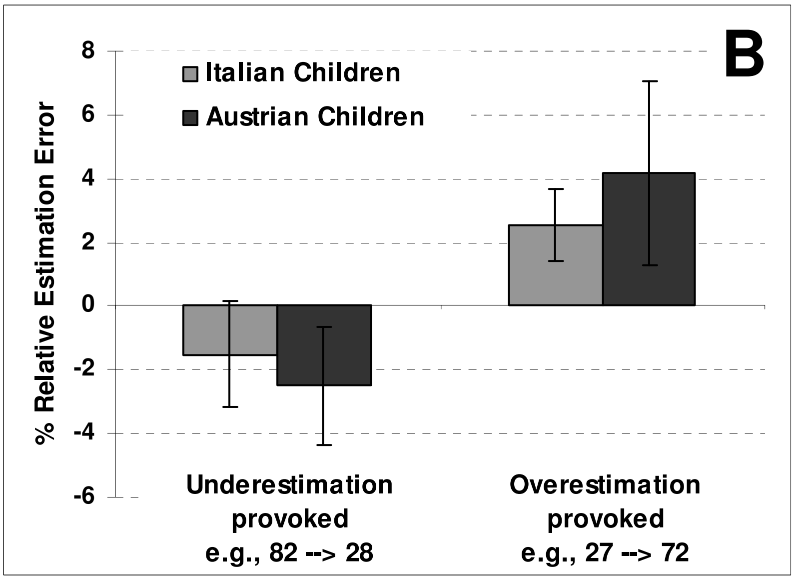
From the notion of the representation of two-digit numbers reflecting the integration of its constituting digits it can be hypothesized that Austrian children might sometimes mix up tens and units of Arabic two-digit numbers, as in the German verbal notation the order of tens and units is reversed compared to the Arabic notation. Misplacements of the digits within the place x value structure of the Arabic number system would then lead to wrong magnitude estimations upon the hypothetic number-line. For testing this hypothesis, items from the 0 - 100 scale were grouped according to the distance relations between their constituting digits (e.g., for 82: 8-2 = 6). For items with a small inter-digit distance (i.e., in our stimulus sample a distance of 2, such as in 64, 35, 75, 53), mixing up the digits should lead to smaller deviations in number-line estimation as compared to items in our stimulus sample with a digit distance larger than 2 (i.e., 47, 95, 27, 82). Accordingly, an ANCOVA on mean absolute error incorporating the within-subject factor interdigit distance (i.e., small vs. large), the between subject factor language (i.e., German- vs. Italian-speaking) and the individual CFT T-scores as covariate revealed a reliable interaction of interdigit distance and language [$F(1,222) = 11.41$, $p < 0.001$]. This indicated that for items with a small interdigit distance no effect of language was present (mean error: 12.65% for Italian and 14.11% for Austrian children [$t(223) = 0.90$, $p = .37$], see Figure 5A) whereas the language effect for items with a large interdigit distance was reliable (mean error: 15.89% for Italian and 21.66% for Austrian children [$t(223) = 2.88$, $p < .01$]). Moreover, in line with the results for all items German-speaking children's estimations were more deviant from the to-be-estimated number than the estimations of the Italian-speaking children as indicated by a reliable main effect of language [14.27% vs. 17.89% estimation error for Italian- and German-speaking children, respectively; $F(1, 222) = 12.28$, $p < .01$]. Furthermore, the significant main effect of interdigit distance [$F(1, 222) = 11.41$, $p < 0.001$] indicated that estimates for items with a small interdigit distance were generally more accurate than estimates for items with a large interdigit distance (18.77% vs. 13.38% error, respectively). Finally, the influence of the covariate (i.e., individual CFT-1 T-values) also reached significance [$F(1, 222) = 4.93$, $p < 0.05$].

More specific under- and/or overestimation in German-speaking children

Directly testing whether the hypothesis that German-speaking children's under-/overshoot should be more pronounced revealed that the difference in relative estimation accuracy between items provoking underestimation and items provoking overestimation was stronger in German- than in Italian-speaking children [$t(224) = 1.70$, $p < .05$, tested one-sided]. As can be observed from Figure 5B this meant that relative under-/overestimations were indeed more pronounced in German-speaking children.

To sum up, we did not find significant performance differences between Italian and Austrian children for items with a small digit distance, whereas reliable differences were present for items with large inter-digit distances. This indicated language-specific differences only for those items where inversion errors would lead to a large deviance between the misunderstood and the actual value of the to-be-estimated number. This interpretation is further corroborated by the finding that German-speaking children exhibited a more pronounced estimation performance difference between items either provoking under- or overestimation of the given number in the expected direction. Taken together, this corroborates our general interpretation that the poorer performance of German-speaking children on the number line estimation task seems to be influenced by the German number word system being less transparent regarding the retention of the place-value structure of the Arabic number system.





<u>Figure 5:</u> The mean estimation error in percent for Italian- and German-speaking children separated for items with a small and large interdigit distance is depicted in Panel A. The relative under-/over-estimation of specific numbers as possibly caused by confusing tens and units is given in Panel B, again separated for the two langusge groups. Error bars indicate 1 SEM.

DISCUSSION

This study set out to examine the role of language in the development of the mental number-line in children. Indeed, we observed two major differences between Italian- and German-speaking first graders. First, the Italian children showed a reliably more accurate estimation performance than the Austrian children despite somewhat lower general cognitive abilities in our sample. Second, on the 0-to-100 scale estimation performance was determined significantly by interdigit distance for German- but not for Italian-speaking children: when processing the inversion property of German number words incorrectly lead to a number deviating largely from the correct number (e.g., 72 instead of the correct 27) number-line estimation was particularly poorer in German- as compared to Italian-speaking children. However, no such language difference was present for items with a small interdigit distance (e.g., "five and thirty" → 53 instead of the correct 35), where incorrect processing of tens and units resulted in a much less deviation error. Finally, the current data for Italian-speaking children corroborated the assumption of first-graders' initially having two separate and linear representations for single and two-digit numbers (Moeller et al., 2009a) rather than one overall logarithmic representation (e.g., Opfer & Siegler, 2007). Actually, the current results for Italian-speaking children were identical to those observed for their German-speaking counterparts as reported by Moeller and colleagues (2009a). In the remainder of this article the implications of these results will be discussed in turn.

Language differences in estimation accuracy

On a general level we observed the number line estimates of Italian-speaking children to be more accurate than the estimates of the German-speaking children on the 0-to-100 scale. Interestingly, this difference did not seem to be driven by differences in age or IQ-test scores since the German-speaking children who performed poorer on the number line task were on average older and scored higher on the CFT-1 than the children of the Italian sample who nevertheless performed better on the number line task[3]. Incorporating the *T*-values of the CFT-1 as a covariate in all analyses did not change this overall result pattern. Moreover, as the first grade curriculum for mathematics is more or less identical in Italy and Austria (both comprising the mastery of numbers up to 20 and their cardinal and ordinal relations as well as first arithmetic procedures within this range, see above) it is unlikely that differences in mathematics education account for the observed performance differences favouring the Italian-speaking children. Thus, age, intelligence-test scores or schooling differences do not

[3] Please note that this difference remained even when taking the age group instead of grade as the reference suggesting that it was not driven by differences in the age of the participants.

seem to be the origin of the performance dissociation between German- and Italian-speaking children. To briefly recapitulate, the German number word system differs from the Italian one as far as for all two-digit numbers above 20 (excluding multiples of 10) the order in which tens and units are spoken is reversed as compared to the order of symbolic digits constituting a given number (e.g., 27 → "siebenundzwanzig" [seven and twenty]). On the other hand, the Italian number word system is fully transparent on this (e.g., 27 → "ventisette" [twenty seven]). To evaluate the influence of this inversion property of the German number word system, the used two-digit items (above 20) were distinguished by the interdigit distance between their constituting digits, thereby, capturing the estimation error resulting from a misconception of tens and units in this number. For instance, when mixing up tens and units of the number 27 and consequently trying to locate 72 on the hypothetic number line, this leads to a larger misplacement upon the mental number-line than mixing up 45 and 54. As unit-decade integration into the place-value structure of the Arabic number system may be more difficult in a language with inversion (cf. Nuerk et al., 2005), German-speaking children should be more likely to may mix up tens and units. Following this argument, we expected general estimation performance to be driven by particular difficulties of the German-speaking children for the items with a large interdigit distance as these were especially affected by inversion errors. Our analyses confirmed the hypothesized language-specific (inversion) effect: German-speaking children's estimates were particularly less accurate for items with a large digit distance, whereas no language difference could be observed for items with a small digit distance for which estimation errors were almost identical. Taken together, these results indicate that estimation accuracy of German-speaking children is particularly poor for items for which inversion plays a major role in the accurate representation of number magnitude. In summary, these findings suggest that the observed language-specific performance difference on the 0-to-100 scale can be attributed to the inversion property of German number words.

Implications for the theoretical understanding of numerical cognition

As outlined above we observed particular language differences in estimation accuracy in the number-line task. In the following paragraph the implications of this finding concerning the theoretical conceptualization of number processing as proposed by the most influential model, i.e., the so-called Triple Code model (Dehaene & Cohen, 1995; 1997; Dehaene, Piazza, Pinel, & Cohen, 2003) shall be addressed. The Triple Code model postulates three distinct representation systems for numerical information that are directly interlinked: a visual system representing the visual identities of the digital input, a verbal system involving number words and arithmetic facts and a quantity system coding numerical magnitude information. Numerical input can directly access all three of the systems dependent on the format in which it is presented. The input information is then transcoded into the appropriate representation for solving the task. This means, depending on task demands one or more of the representation systems are involved. From the beginning, the representation of number magnitude in the quantity system was assumed to be organized as an internal MNL with numbers being arranged in ascending order from left to right upon this MNL (cf. Dehaene et al., 1993; Dehaene & Cohen, 1995). Consequently, number magnitude seems to be associated with physical space (see Bueti & Walsh, 2009; de Hevia, Vallar, & Girelli, 2008; Umiltà, Priftis, & Zorzi, 2009 for recent reviews on different aspects of spatial-numerical associations; Wood, Willmes, Nuerk, & Fischer, 2008; for a review and a meta-analysis). In this context, the number-line task intends to measure the nature and stability of (linear) spatial-numerical associations.

How can language properties exert an influence on the spatial representation of number magnitude? According to the model, at least three processing steps would be required for number-line estimation: (i) access to the numbers to-be-estimated through the visual Arabic system; (ii) transmission to the spatial representation system where magnitude is processed analogously upon a mental number-line; and (iii) direct retrieval of the output from this system. In this case no language-specific effects would be predicted. In order to explain the influence of language properties on the spatial representation system an interaction with the verbal system needs to be assumed. Such an interaction has been proposed by Nuerk et al. (2005) who reported language effects in a task as basic as magnitude comparison. Their results indicated that the verbal representation of numbers interacts with the magnitude representation, even for the case that numbers are presented in Arabic digital notation. Despite the fact that magnitude comparisons should be possible to perform without involving the verbal representations of the respective numbers, Nuerk et al. (2005) suggested that the Triple Code model does not per se exclude the possibility of a co-activation of the verbal representation even when it is not of primary relevance for the task at hand.

Considering the current study, the data clearly corroborate the assumption that the verbal number representation influences the number line task, which is usually assumed to exclusively assess the spatial representation of number magnitude. A possible explanation of this influence in the present study may stem from the development of the neural circuitry underlying numerical cognition. Recent functional MRI studies indicated that the cortex areas associated with the three representational codes of numerical information are much less distinct in children than in adults (Kaufmann et al., 2006; Kaufmann et al., 2008; Rivera, Reiss, Eckert, & Menon, 2005). This means, for the same numerical task children activate a more wide-spread neural network possibly including activation of several representations in parallel, even when they are not primarily needed for solving the task. Thus, in number-line estimation the to-be-estimated numbers seem to co-activate their verbal word frame. In German-speaking children inverted two-digit number words on the verbal and non-inverted digital Arabic representation may both be elicited and thus compete for use. When the influence of the verbal representation is larger for one item it would necessarily determine a misplacement of the item on the hypothetical number-line. As such an influence would be especially detrimental in items with a large interdigit distance, language differences should be most pronounced in these items – as they were in the current study.

It may be important to note that these data do not necessarily imply that the spatial representation of numbers is verbal. Rather, they could imply both (i) that the access to the spatial representation of number magnitude may be verbally mediated or (ii) that the spatial representation of numbers itself is moderated by verbal co-activations. In research on spatial number representations in neglect patients, this point has received much interest recently (Priftis et al., 2006; Umiltà et al., 2009). Neglect patients do not only have problems with the spatial representation of number magnitude per se, but also with the integration of tens and units in the place-value system in the neglected portion of the number line (cf. Hoeckner et al., 2008). Future cross-cultural studies with neglect patients may help to evaluate the question whether the magnitude representation itself (including place-value attributes) or the access to it is influenced by language differences.

In summary, the observed language-specific effects on the number-line task can best be accounted for by assuming a less specific activation of the distinct numerical representations in children as compared to adults. In our view, the Triple Code model (Dehaene & Cohen, 1995; 1997) can only be applied to children data, when a more integrative interaction of its single systems is assumed during the acquisition of number processing skills (see also Cohen & Dehaene, 2000, for similar suggestions to account for adult patient behavior).

Logarithmic vs. two-linear representation of numbers up to 100

In a recent study Moeller and colleagues (2009a) showed that the mental representation of numbers up to 100 in German-speaking first graders may not only be accounted for by a logarithmic model (e.g., Siegler & Booth, 2004) or a two-linear model with a breakpoint associated with children's familiarity with numbers, but alternatively by a two-linear model with one linear representation of one-digit numbers and another linear representation of two-digit numbers. In the data of Moeller and colleagues (2009a) both models captured the data well (i.e., the logarithmic as well as the two-linear with a fixed breakpoint at 10). However, in direct comparison and some critical tests the two-linear model outperformed the logarithmic model reliably in terms of descriptive adequacy (i.e., R^2) in a German-speaking children sample. From this finding, Moeller et al. (2009a) conclude that the spatial coding of numerical magnitude along the MNL may not develop from logarithmic to linear with increasing age and experience as previously suggested by Siegler and colleagues (e.g., Opfer & Siegler, 2007; Siegler & Booth, 2004). Instead, the developmental change may involve the mastery of the place-value structure of the Arabic number system. As observed by Moeller and co-workers (2009) as well as by many others (e.g., Siegler & Opfer, 2003; Opfer & Siegler, 2007) children seem to systematically overestimate the space reflecting the position of one-digit numbers towards the right in the beginning (i.e., locating 7 at the actual position of approximately 30). However, on average they also adhere to the ordinal order of the to-be-estimated numbers. Moeller et al. (2009a) interpreted this as indicating that children already seem to know that, for instance, 40 is somehow larger than 4; nevertheless, they are not yet capable of correctly representing the ratio of this "larger" complying with the base-10 place-value structure of the Arabic number system (i.e., the distance between 0 and 40 is 10 times as large as the distance between 0 and 4). When the "somehow" in above sentence means not 10 times larger, but 2 or 3 times larger, this could potentially produce data patterns which are well fitted by a logarithmic fitting although the underlying data may not be logarithmic[4]. Thus, it may not be a logarithmic to linear change of children's magnitude

[4] Please note that this argument holds for both directions. Data produced by a logarithmic model can also be well fitted by a two-linear model, at least when no fixed break points around 10 is assumed as in the model of Ebersbach et al. (2008). Thus, descriptive adequacy alone cannot prove that the underlying data are also produced by either model fitting the data (cf. Myung & Pitt, 1997, see also Appendix A for further discussion of this point).

representation but rather progress in children's understanding of the place-value structure that underlies performance improvements with age and experience.

Moeller and colleagues (2009a) presented evidence indicating a better fit of the two-linear compared to the logarithmic model in various indices for German-speaking children only. However, it is known that the intransparent German number word system, in particular its inversion property, has a reliable influence on children's numerical development (cf. Zuber et al., 2009). Therefore, it was still to be shown that the results reported by Moeller and co-workers (2009a) were not determined by this peculiarity of the German number word system. In the present study, we evaluated the spatial representation of number magnitude in 107 Italian-speaking first graders. As there is no inversion of tens and units in the Italian number word system, replicating the better two-linear as compared to a logarithmic fit would corroborate the notion of children's mastery of the place-value structure of the Arabic number system to determine children's spatial representation of number magnitude irrespective of differences in the number word systems.

The current data support the latter assumption. The results of the Italian-speaking children mirrored those of the German-speaking ones in each and every aspect: (i) comparing logarithmic to two-linear fittings on an individual basis showed that the two-linear model provided a reliably better fit; (ii) the vast majority of children exhibited a breakpoint around ten, and (iii) the slope of the linear segment representing one-digit numbers was negatively correlated to the slope of the segment for two-digit numbers. Taken together, the results for the Italian-speaking children were identical to those previously observed for German-speaking children (cf. Moeller et al., 2009a). In our view, this data indicate that the mastery of the base-10 place-value structure of the Arabic number system to be a corner stone in the development of the MNL in different cultures. Even more, the present results imply that the general way in which children learn to represent numbers complying with the base-10 property of the place-value structure of the Arabic number system (i.e., the distance between 0 and 40 is 10 times as large as the distance between 0 and 4) is not influenced by language differences. Finally, Booth and Siegler (2008) also showed that performance in the number line task is associated with actual mathematics achievement and in particular the ability to solve unlearned problems. To substantiate these findings we evaluated whether the size of the individual mean estimation error in first grade predicted the mathematics mark two years later by the end of grade three for a subsample of 50 of the German-speaking children. And indeed, in line with the results of Booth and Siegler (2008) we observed that third grade mathematics mark became better, the smaller the number line estimation error was in first grade (r = .40; n = 50). Synced with above findings suggesting two separate representations of single- and two-digit numbers one could as well assume that as an alternative account to the logarithmic-to-linear shift, place-value understanding serves as a building block for later arithmetic competencies.

SUMMARY AND CONCLUSIONS

This study set off to gain new insights in the development of the spatial representation of number magnitude and the role of language properties on this representation system by evaluating performance of Italian- and German-speaking first grade children in a number-line estimation task. The latter languages were chosen as they differ in the organization of their number word systems. While the order of tens and units in German number words is reversed as compared to the digital notation (i.e., 27 → seven and twenty), no such inversion is present in Italian. Recently, the inversion property of German number words was shown to pose specific problems to children when acquiring the Arabic notation of multi-digit numbers (Zuber et al., 2009). The present data are informative in two respects reflecting language independent as well as language dependent development:

First, it was observed that comparable to previous results for German-speaking children estimation performance of Italian-speaking children was accounted for best by a two-linear model as well. In line with Moeller et al. (2009a) we argue that this corroborates the notion that performance improvements with age and experience may be driven in by progress in the children's understanding of the base-10 place-value structure of the Arabic number system. In this context, the current results indicate that the process of mastering the base-10 property of the place-value structure may be language invariant as we observed no differences between German- and Italian-speaking children.

On the other hand, despite the fact that the underlying processes may be the same in both languages language differences in the number word system nevertheless influenced the execution of these processes. Estimation performance for two-digit numbers was particularly worse in German-speaking children when confusing tens and units in integrating the individual digits into the overall number (due to their inverted order in German number words) results in large differences between the actual and the misunderstood number (i.e., seven and twenty → 72 instead of 27). This indicated that although German-speaking children refer to the same underlying process of integrating tens and units into one coherent representation of a two-digit number, the fact that the input to this process is diverse (verbal vs. digital notation) takes its toll on estimation accuracy.

To conclude, we want to emphasize the importance of translingual studies for investigating the development of numerical cognition. The present study indicated that there are language-independent as well as language-dependent influences. However, in both cases general as well as language specific processing within a language could only be identified with the reference to at least one other language. Without such a direct contrast between two languages with e.g., different number word systems it is almost impossible to dissociate language specificities from more general developments of numerical competencies. In this vein, future translinguistic studies provide the possibility to better understand which aspects of numerical development follow rather language invariant trajectories as well as which language-specific properties influence these trajectories and how this influence looks like.

APPENDIX A

A remark on how to evaluate the logarithmic vs. two-linear issue

There are different ways of evaluating whether the logarithmic or the two-linear model provide a better fit of the empirical data. In the following, these different approaches shall be summarized. Generally, (i) approaches averaging data across participants can be distinguished from (ii) approaches contrasting individual performance measures.

Averaging across participants (fixed effects analysis):

(a) After averaging the estimates for each to-be-located number across all participants a comparison may be of interest which of the two models (logarithmic vs. two-linear) provides a better fit of the empirical data (e.g., in terms of adjusted R^2). This can be done for instance by a stepwise multiple regression analysis in which both a linear as well as a logarithmic predictor is entered. By the logic of the stepwise regression analysis the predictor which accounts for the largest part of the variance is included in the model first. Any further predictor will only be incorporated into the model when it adds significantly to the variance explained by the final regression model. Thereby, it would be possible to identify which of the two predictors (i.e., linear or logarithmic) is incorporated into the regression model first and to evaluate whether the inclusion of the other predictor would add reliably to the explanatory power of the model or not. When running this stepwise regression analysis on mean estimates of all Italian-speaking children it could be observed that only the two-linear predictor was considered in the final regression model while the logarithmic predictor was not incorporated [R = .99, adj. R^2 = .97, $F(1, 16) = 553.59$, $p < .001$, $b_{\text{two-lin}} = .99$, $p < .001$, $b_{\log} = -.35$, p = .18]. This observation was substantiated by the results of two forward regression analyses. When entering the logarithmic and the two-linear predictor successively the logarithmic predictor on its own was found to be a reliable predictor of estimation performance [$b = .96$, $p < .001$]. However, when entering the two-linear predictor the logarithmic predictor was no longer a reliable predictor of estimation performance [R = .99, adj. R^2 = .97, F(2, 15) = 295.11, p < .001; see Table A]. On the other hand, when starting the regression analysis with the two-linear predictor and then adding the logarithmic predictor the latter did not account for a reliable portion of additional variance [R = .99, adj. R^2 = .97, F(1, 16) = 553.59, p < .001, $b_{\text{two-lin}}$ = .99, p < .001, $b_{\log}$ = - .35, p = .18]. Please note that the results for the German-speaking sample were identical.

<u>Table A:</u> *Results of the first logarithmic then two-linear forward regression analysis for Italian-speaking children*

Predictor	***B***	**Standardized *b***	**Change in R^2**	***t***	***p***
Logarithmic	- .36	- .35	.93	1.42	.18
Two-linear	1.35	1.33	.05	5.43	.001

(b) Identifying the optimal break-point in a two-linear regression may provide additional information. Only if the optimal break-point falls near the theoretically proposed break-point the two-linear model seems appropriate. As can be seen from Figure 2 adjusted R^2 for the two-linear model is descriptively larger than that for the logarithmic model. Furthermore, the optimal break-point of 9.57 for the Italian-speaking as well as 11.33 for the German-speaking sample was indeed very close to the hypothesized break-point of 10.

(c) Finally, running separate analyses for one- and two-digit numbers may also be of particular interest for two main reasons. First, when assuming each of these two representations to be linear with a fixed breakpoint at 10 instead of an overall logarithmic representation of number magnitude such analyses reflect a crucial test of the predictions of the two-linear model. On the other hand, when assuming an overall logarithmic representation not differentiating between single- and two-digit numbers separate analyses for each of these intervals should nevertheless reveal that the estimates are fitted best by a logarithmic function.

Splitting the overall interval in two should do nothing to the validity of the logarithmic model. This may be of particular interest for the segment representing two-digit numbers as the proposition of a linear representation of all two-digit numbers seems to be a strong claim against the background of the problem size effect, which should be tested prior to any further analysis. Given the case there is no problem size effect for the two-digit numbers (or e.g., the larger half of the items), interpreting the relatively good fit of a logarithmic function to indicate a logarithmic magnitude representation may be premature as there might actually be no explicit representation of the magnitude of these items. Instead, this might reflect the case of intact magnitude representation up to a given quantity (e.g., 10) with each magnitude above this point corresponding to a rather fuzzy representation of "many". Please note that the assumption of an overall magnitude representation as claimed by the logarithmic model is violated by such a data pattern. On the other hand, a two-linear model as proposed by Ebersbach et al. (2008) would be able to account for the data pattern in a more appropriate way. Possible confounds to a logarithmic interpretation of this kind may for instance be inherent in the data pattern reported by Opfer, Thompson, & Furlong (in press, Figure 5, p. 8) interpreted to indicate a logarithmic magnitude representation of even preschoolers (see also Brysbeart, 1995; Muldoon, Simms, Towse, Burns, & Yue, this issue for comparable data patterns). Second, by running linear and logarithmic analyses separate for one- and two-digit numbers as testing specific model predictions of the two-linear model with a breakpoint at 10 no special adjustment for the degrees of freedom of the two models is necessary as these are identical (i.e., linear: $y = a \cdot x + b$; logarithmic: $y = a \cdot \log(x) + b$).

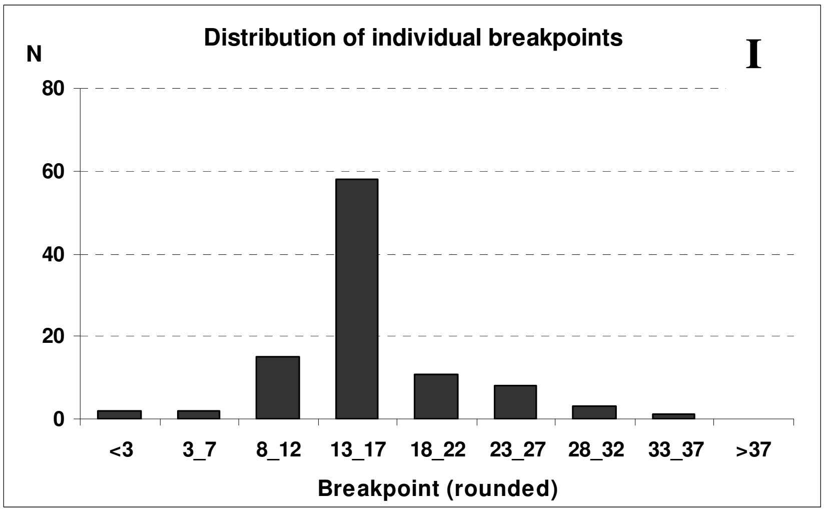
Fitting individual participants (random effects analysis):

(a) When intending to evaluate model adequacy on a more individual basis a measure of model fit (e.g., adjusted R^2) can also be calculated for both the logarithmic as well as the two-linear model for each participant individually. Afterwards the two matrices are directly contrasted. In the current study it was observed that the model fit of the two-linear model was reliably better than that of the logarithmic model for Italian-speaking children (see also Moeller et al., 2009a for a more detailed description of identical results for the German-speaking sample).

(b) Alternatively, one may also be interested in the results of separate analyses for one- and two-digit numbers as described above on an individual bases.

(c) It is also possible to compute the optimal break-point of the two-linear model for each individual participant and then evaluate the distribution of these break-points. Only if the optimal break-points of the vast majority of participants fall close to the hypothetisized break-point the two-linear model should be considered. Again, this was given in the present data set: the majority of Italian-speaking participants had an optimal break-point around 10 (see also Moeller et al., 2009a for similar results concerning the German-speaking participants).
However, as already pointed out in the introduction even in a situation where the data undoubtedly follow a two-linear function, a logarithmic model would provide a very good model fit and vice versa. Therefore, we evaluated this issue by comparing the distributions of individual breakpoints of the empirical data and 100 simulated participants produced by a 100% logarithmic model. Figure A, Panel I depicts the distribution of breakpoints as observed for the simulated data whereas the individual breakpoints empirically found for both German- as well as Italian-speaking children are to be found in Figure A, Panels II and III, respectively. As can be observed from these figures the pattern differs considerably: For the simulated data the vast majority of breakpoints was found to be located between 13 and 17 with the median at 14.37 and the mean being very close at 15.39. Contrarily the majority of breakpoints for the empirical data was found to be located between 8 and 12 (median at 9.8 and 11.8 for Italian- and German-speaking children, respectively; Means: 21.2 and 25.3, respectively) and thus

much closer to the proposed breakpoint at 10 reflecting two different representations for single- and two-digit numbers. On the other hand, we are not aware of any theoretical assumption suggesting differing representations of numbers below and above 14. Another interesting finding is that of the 100 individual breakpoints of the simulated data only 4 were to be found below 10 whereas the number of breakpoints below 10 was higher for the empirical data (German-speaking: 34; Italian-speaking: 42) of which the majority was located on 8 and 9 (German-speaking: 21; Italian-speaking: 26).



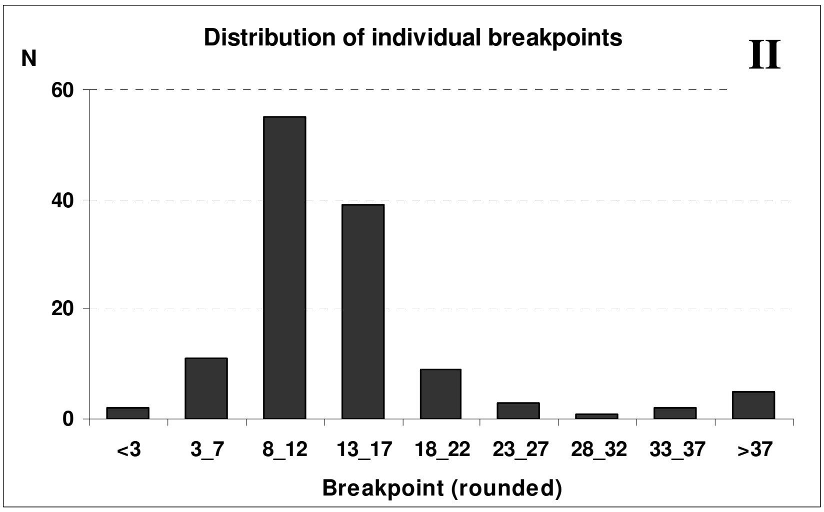
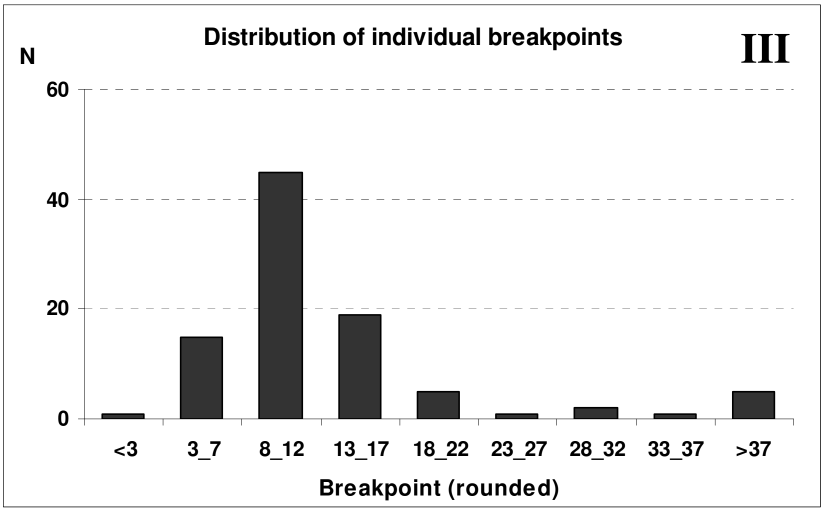


Figure A: Distribution of breakpoints for the data simulated by a logarithmic model (Panel I), German-speaking (Panel II) and Italian-speaking (Panel III) children



(d) Finally, a correlation analysis of the individual slopes of the two linear or two logarithmic segments for one- and two-digit numbers may be informative. Under the assumption of the two-linear model suggested by Moeller et al. (2009a) the slopes of the two linear segments of the two-linear model should be negatively correlated: the steeper the slope of the segment for single digit numbers the further to the right are one-digit numbers placed on the hypothetical number line. Thereby, all two-digit numbers have to be located on a very limited part of the number line leading to a relatively flat slope for these two-digit numbers. On the other hand, the slopes of two logarithmic segments meeting at a fixed breakpoint should be positively correlated. Thereby, evaluating the correlation of the slopes of the two linear segments for one- and two-digit numbers is a critical test for the applicability of the two-linear model and the two separate representations of one-and two-digit numbers it reflects. For both the Italian- as well as the German-speaking sample the correlation between the slopes of the two linear segments was significantly negative.

It is obvious that both approaches follow the same logic either across items (see above) or across participants. However, in the latter approach model fit is assessed on an individual level, thereby, taking into account possible individual differences in development which are treated as measurement error in the item-based approach. Therefore, we think that the individual approach may be more suited to investigate developmental trajectories of the spatial dimension of number magnitude representation.

Addressing the problem of possible model overfitting

Finally, a general issue in science regards the question of model overfitting in the way of evaluating whether it is signal which is modeled by a given model or rather signal plus noise. In the latter case, one runs the risk of capitalizing on measurement error due to a more complex model with a higher number of degrees of freedom (see Myung & Pitt, 1997; Pitt, 1999, Pitt, Myung, & Zhang, 2002 for a discussion of this point). Transferred to the current study this taps on the different number of free parameters between the logarithmic and the two-linear model. While the logarithmic model [based on a function such as $y = a \log(x) + b$] has the two free parameters a and b, a two-linear model without a fixed breakpoint has four free parameters (a_1, a_2 and b_1, b_2) two for each linear part of the model (i.e., $y = a_1 x + b_1$ for the first part and $y = a_2 x + b_2$ for the second part). Yet, when using a fixed breakpoint (as assumed for the current two-linear model) the number of free parameters comes down to three. When the first linear segment is described by the function $y = a_1 x + b_1$ with two free parameters and the constraint of ending at a fixed breakpoint (e.g., 10) it defines a fixed point at this endpoint which the second linear part of the model has to cut. Thereby, the second part of the two-linear model with a fixed breakpoint is described sufficiently by one further free parameter, be it either constant or slope. As the second linear element has to cut a fixed point at the breakpoint its slope is fixed when the constant is estimated while on the other hand the constant is fixed when the slope is estimated. Taken together, the proposed two-linear model with fixed breakpoint at 10 has exactly *one* free parameter more than the logarithmic model. Nevertheless, this could make the difference. To account for the higher number of free parameters in our model as compared to both the simple linear as well as the logarithmic model R^2 values were adjusted by the number of free parameters of either model prior to any analyses. However, in the following, the results of two additional analyses will be presented addressing the problem of possible overfitting of the data by the more complex two-linear model.

In a first approach it was evaluated in how far the results observed depended on the inclusion of one particular data point (i.e., to-be-estimated number) to examine over-fitting of the data. When iteratively omitting one item from the analysis results in considerable changes of each of the models' performance this indicates that the overall result is highly dependable on the inclusion of particular data points. On the other hand, when results do not change considerably this argues for a high stability of the overall result indicating that it is rather invariant to the exclusion of specific items. A respective analysis for the current data showed that the results for both models were very stable (logarithmic: mean adjusted R^2 = .68, SEM < 0.001; two-linear: mean adjusted R^2 = .82, SEM < 0.001; mean number of participants with an optimal break-point around 10: 62.72, SEM = 0.87; mean slope of single digit segment: 3.93, SEM < 0.01; mean slope of two-digit segment: 0.32, SEM < 0.001). More interestingly, the two-linear model outperformed the logarithmic one reliably in terms of adjusted R^2 no matter which data point was omitted from the analysis (all $t > 10.23$, all $p < .001$). Note that this stability is true for a logarithmic model and for a two-linear model with a fixed break point. A model with a variable break point could be potentially more flexible and thus more variable when single data points are omitted.

Another way to address the point of possible overfitting of the two-linear model is to evaluate in how far a two-linear model accounts for data produced by a logarithmic model and vice versa. Generally, data produced by logarithmic model (including some random noise) must not be accounted for better by two-linear fitting as compared to logarithmic fitting. When data produced by a logarithmic model is accounted for better by two-linear fitting than by logarithmic fitting this would suggest overfitting of the data by the two-linear model.

Contrarily, data produced by a two-linear model (again including some random noise) must not be fitted better by a logarithmic model. To test these predictions we simulated the data of 100 participants using both a logarithmic as well as a two-linear model incorporating a Gaussian random error term and fitted the resulting data in a logarithmic and a two-linear regression analysis for each of these modelled data sets. Subsequently, adjusted R^2 resulting from either regression were compared using paired t-tests. The results of the t-tests showed that logarithmic fitting did account for reliably more variance of the data produced by a logarithmic model than did two-linear fitting [$t(99) = 1.99$, $p < .05$; $R^2_{\text{log}} = .94$ vs. $R^2_{\text{two-lin}} = .93$]. On the other hand, two-linear fitting resulted in a higher adjusted R^2 than did logarithmic fitting for the data produced by a two-linear model [$t(99) = 10.28$, $p < .001$; $R^2_{\text{log}} = .87$ vs. $R^2_{\text{two-lin}} = .93$]. Apart from the fact that all analyses were run on R^2 values adjusted for differing numbers of free parameters this finding again argues against an overfitting of the empirical data by the two-linear model.

Study 4

Early place-value understanding as a precursor for later arithmetic performance

–

A longitudinal study on numerical development

ABSTRACT

It is assumed that basic numerical competencies serve as a cognitive scaffold for the development of more complex arithmetic skills. The current study aims to take a first step in evaluating this interrelation in a longitudinal approach involving typically developing children. It was investigated whether first graders' performance in basic numerical tasks in general as well as specific cognitive processes involved in such tasks (e.g., place-value understanding) reliably predicted their performance in a complex addition task in third grade.

The results indicated that not only overall performance in basic numerical tasks influenced later arithmetic achievement. More particularly, an early understanding of the place-value structure of the Arabic number system was observed to be the most reliable predictor for specific aspects of arithmetic performance. Therefore, we suggest that a representation of the place-value structure of the Arabic number system should be considered in current models of numerical development. Implications of the role of basic numerical competencies for the acquisition of complex arithmetic and its impairment are discussed.

INTRODUCTION

Dealing with (multi-digit) numbers is maybe one of the most important abilities learned at school as it has a wide range of everyday applications. As a consequence, impaired numeracy poses substantial problems for children's educational development (Bynner & Parsons, 1997) as well as post-educational experiences throughout life (Parsons & Bynner, 1997; 2005). Thus, given the importance of these skills understanding their developmental trajectories and their cognitive underpinnings seems crucial.

Indeed, in recent years the research interest in calculation processes, calculation strategies as well as their underlying cognitive architecture increased steadily with a clear focus on investigating addition (e.g., Adams & Hitch, 1997). Addition is the first arithmetic operation taught at school. Moreover, according to Ashcraft (1992) successful mastery of addition rules and procedures constitutes an essential building block for the development of more advanced mathematical skills (e.g., multiplication; Cooney, Swanson, & Ladd, 1988).

When developmental trajectories of more advanced mathematical skills are investigated, it is important to know which more basic cognitive characteristics underlie such advanced mathematical skills. Recently, there has been intriguing research investigating (multi-digit) addition identifying different characteristics of addition problems moderating task difficulty such as problem size (Deschuyteneer, De Rammelaere, & Fias, 2005; Imbo & Vandierendonck, 2007a; Stanescu-Cosson et al., 2000), carry over, (Klein, Nuerk, Wood, Knops, & Willmes, 2009; Imbo, Vandierendonck, & De Rammelaere, 2007a), solution strategies depending on these characteristics (Roussel, Fayol, & Barrouillet, 2002; Imbo, Vandierendonck, & Rosseel, 2007b) and their relation to working memory (e.g., De Smedt et al., 2009a; Imbo, Vandierendonck, & Vergauwe, 2007c).

On the other hand, knowledge of which basic numerical precursor competencies predict the acquisition of addition skills reliably is still rather patchy. Although many authors have claimed that basic numerical knowledge serves as an important precursor for later addition performance and arithmetic achievement in general (e.g., Butterworth, 2005; Dehaene, 1997), systematic investigation of this relationship has only began recently. In a cross-sectional approach, Holloway and Ansari (2009) found children's early ability to compare symbolically coded numerical magnitudes to be a predictor of their mathematical skills. The authors assessed children's ability to compare symbolic (i.e., digital notation) and non-symbolic (i.e., dot patterns) magnitudes. They found that only the distance effect observed in symbolic magnitude comparison reliably correlated with their performance in a standardized mathematical test. These findings provide direct empirical evidence that complex calculation performance relies upon basic numerical skills, such as magnitude understanding (see also Kaufmann, Handl, & Thoeny, 2003). However, apart from this important observation, the study is innovative for another aspect: Unlike the majority of previous studies (e.g., Booth & Siegler, 2008) Holloway and Ansari (2009) did not index one (numerical) representation by one numerical task (henceforth: *task approach*; e.g., indexing the quality of magnitude representation by overall error rate in a number comparison task). Rather, Holloway and Ansari (2009) used a *specific numerical effect*, i.e., the numerical distance effect, as a more stringent index for the underlying representation, i.e., number magnitude representation (henceforth: *effect approach*).

As the distinction between task approach and effect approach is crucial for the current study and beyond, we will elaborate on that distinction a bit more. Consider the following as an example for the *task approach*: Dehaene and Cohen (1997) concluded that the magnitude representation of their patient MAR was impaired as he was moderately to severely impaired (i.e., exhibiting an abnormally high error rate) in a number of quantitative numerical tasks including magnitude comparison and number bisection amongst others. Thus, a specific task (in this case number magnitude comparison) is used to index a specific representation (in this case number magnitude representation). At first sight, this may seem convincing. However, it has been shown that other representations influence task performance in such tasks as well. For instance, Nuerk, Weger, and Willmes (2005) have shown non-magnitude-related language effects in the magnitude comparison task whereas Nuerk, Geppert, van Herten, and Willmes (2002, see also Korvorst & Damian, 2008; Wood et al., 2008) observed reliable non-magnitude effects of parity and multiplication fact knowledge in the number bisection task.

On the other hand, following the *effect approach* an underlying cognitive representation (e.g., magnitude representation) is reflected by the size of a specific effect associated with this representation (e.g., the numerical distance effect). As the distance effect is one of the most robust effects in numerical cognition (observed even in non-human primates, e.g., Nieder, Diester, & Tudusciuc, 2006) its disappearance, reversion, or inflation can be used to indicate impairments of the underlying number magnitude representation. In this vein, Delazer and colleagues (Delazer, Karner, Zamarian, Donnemiller, & Benke, 2006) observed an increased distance effect to indicate impaired number magnitude processing capabilities (despite the fact that overall error rate in a magnitude comparison task did not seem conspicuous). Thereby, the effect approach allows for a more fine-grained evaluation of performance even in situations where overall performance measures (e.g., overall error rates) may be too unsubtle and thus undifferentiated. Therefore, focusing on a specific numerical effect within a task instead of overall performance in this task may provide additional information when investigating mastery of a particular underlying representation (see also Hoeckner et al., 2008; Korvorst & Damian, 2008; Nuerk et al., 2002; Wood et al., 2008).

Such an effect approach has already been successfully applied to the domain of numerical development by Holloway and Ansari (2009). They observed that children who exhibited a comparatively larger symbolic distance effect (i.e., more errors when comparing two numbers close to each other than two numbers further apart, e.g., 3_4 vs. 1_6) scored lower on a standardized mathematics achievement test (see also De Smedt, Verschaffel, & Ghesquière, 2009a; Landerl & Kölle, 2009 for similar results).

The current study follows up and extends the effect approach used in the innovative study of Holloway and Ansari (2009). It aims at extending the study of Holloway and Ansari (2009) in four aspects: (i) Holloway and Ansari (2009) used an effect approach only on the side of the predictor variable (i.e., the symbolic distance effect), while we will use numerical effects as the predictor and the dependent variable. Thus, not only the representation predicting later performance is measured within-task, but also the processes and representations indexed by the criterion variables are measured by specific numerical effects in addition to overall performance measures. (ii) While Holloway and Ansari (2009) tested a cross-sectional sample of children, the present study presents longitudinal data on the developmental trajectories from basic numerical competencies (first grade) to later arithmetic performance (third grade). (iii) The stimulus sets used to assess basic numerical competencies in this study (i.e., transcoding and magnitude comparison) were almost entirely composed of at least two-digit numbers while single-digit numbers were employed by Holloway and Ansari (2009). This directly relates to the last point: (iv) As Holloway and Ansari (2009) only used single-digit stimuli they could not address the issue of place-value understanding as an important precursor of later arithmetic ability in their study. Contrarily, the issue of mastery of the place-value structure of the Arabic number system will be the central question of the current study. Therefore, the relevance of the place-value concept as well as children's difficulties in acquiring this concept will be reviewed in the remainder of this introduction.

The place-value structure of the Arabic number system

Obviously, complex (multi-digit) arithmetic does not only require an activation of number magnitude (e.g., Dehaene & Cohen, 1995; Holloway & Ansari, 2009), but also requires at least basic understanding of the place-value structure of the Arabic number system (Moeller, Pixner, Kaufmann, & Nuerk, 2009a; Nuerk, Kaufmann, Zoppoth, & Willmes, 2004a; Nuerk, Graf, & Willmes, 2006; Nuerk & Willmes, 2005, for a review). The Arabic number system as formalized by its 10 digits (0, 1, 2, 3, 4, 5, 6, 7, 8, and 9) is formally simple because assembling a number by its constituting digits follows the so called place-value principle (Zhang & Norman, 1995). The place-value concept signifies that the value of a digit is defined by its position within the sequence of digits constituting the respective number (Zhang & Norman, 1995): Starting from the rightmost digit the power of 10 is increasing by one with each step to the left (e.g., 368 = $\{3\}\ 10^2\ \{6\}\ 10^1\ \{8\}\ 10^0$; cf. McCloskey, 1992). Taking this into account, complex addition requires quite elaborate understanding of the place-value concept. For instance, consider the problem 36 + 57. Here, unit and decade digits need to be manipulated and arranged correctly to obtain the result 93, i.e., considering the carry, overwriting of zeros: 90 → 93. Thus, more advanced mathematical skills might not only be predicted by a general activation of magnitude representations, but might, in particular, rely on the understanding of the place-value system and the integration of constituting digits in this system as well (Kaufmann & Nuerk, 2007; Nuerk et al., 2004a; Nuerk et al., 2006).

In what follows children's difficulties in the acquisition of the understanding of the place-value system will be briefly sketched and the possible impact on the specific characteristics of complex addition problems will be outlined to allow for a deduction of specific longitudinal hypotheses later.

Understanding the Arabic place-value concept

Several studies indicated that children experience specific difficulties acquiring this principle. One of these difficulties is reflected by problems in mastering the correspondence between verbal number words and the place-value principles for Arabic digits (e.g., Camos, 2008; Power & Dal Martello, 1990; 1997; Zuber, Pixner, Moeller, & Nuerk, 2009). These problems seem to be mainly driven by the fact that in most Western languages the verbal number-word system is not structured along place-value constraints (cf. Comrie, 2005). This argument is corroborated by the important results of Miura and colleagues (1994) who consistently found that Asian children (i.e., Chinese, Japanese, and Korean) experiencing a totally regular and transparent number word system (42 → four ten two) outperformed their western counterparts (i.e., France, Sweden, and the U.S.) in recoding given numbers into non-symbolic sets of tens and units blocks (see also Miura & Okamoto, 1989; Miura, 1987). In this vein, the verbal number word system in most western languages is much more complex than the symbolic Arabic number system as it is organised by different classes of number words: (i) units (0 to 9), (ii) decades (10, 20, etc.), (iii) teens (13, 14, etc.), (iv) hundreds, thousands, etc. For the verbalization of multi-digit numbers these number word classes are combined by two syntactic rules, i.e., multiplicative composition as well as additive composition (involving the requirement to overwrite zeros). For instance, three hundred eighty-nine builds upon a product relationship of 3 x 100 and a sum relationship of 300 + 89 → 389 (cf. Power & Dal Martello, 1990). Typically, when children are required to transcode verbal number words (e.g., twenty-five) into the corresponding string of Arabic digits (i.e., 25) or vice versa, their errors are almost exclusively related to the understanding of these syntactic principles (Barrouillet, Camos, Perruchet, & Seron, 2004; Camos, 2008; Power & Dal Martello, 1990; 1997; Seron & Fayol, 1994; Zuber et al., 2009).

To illustrate this, consider a child who was dictated 124 but wrote 10024 or 1024. Both of these errors reflect a misconception of the base-10 place-value system; in particular the necessity to overwrite zeros (Power & Dal Martello, 1990). However, the syntactic structure of number words differ between languages as for instance in German unit and decade digit are spoken in reversed order (43 is spoken as *three and forty*). This inversion property of the German language has only recently been identified to pose a major obstacle for transcoding performance. In this vein, Zuber and colleagues (2009) assessed transcoding performance in 130 German-speaking first graders who had to write down 64 numbers that were dictated to them. The authors observed that, besides other common syntactic errors (e.g., 124 → 10024, see Deloche & Seron, 1982; Power & Dal Matello, 1990 for a taxonomy of transcoding errors), about one half of the German speaking children's transcoding errors were related to inversion (e.g., forty-three – spoken as three and forty in German – was written as 34). Thus, when number words are not transparent in regard of the place-value system, successful integration of tens and units into the place-value system becomes even more difficult. Taken together, these findings nicely illustrate, that especially for multi-digit numbers transcoding not transparent notational formats poses specific difficulties on children as it requires elaborate understanding and application of rules in reference to the place-value structure of the Arabic number system. (e.g., Proios, Weniger, & Willmes, 2002; Zuber et al., 2009).

Furthermore, children's specific difficulties in the acquisition of the place-value concept do not only become evident from their transcoding errors. Even in tasks as basic as two-digit number magnitude comparison, children are need to understand and more or less automatically apply place-value rules. For example, when comparing 42 and 57 tens and units need to be put in the correct bins (e.g., Nuerk, Weger, & Willmes, 2001), meaning that one has to segregate which digits represent the to-be-compared decades (i.e., 4 < 5) and which the units (i.e., 2 < 7). Such unit-decade identification within the place-value system is especially important when the presented number pair is unit-decade incompatible; meaning that the unit digit of the smaller number is larger than the unit digit of the larger number (e.g., 47_62, 4 < 6, but 7 > 2). Above chance performance in incompatible trials requires necessarily – at least some implicit – place-value understanding (see Table 1; for examples and illustration). Furthermore, it has to be noticed that even for children and adults with an elaborate place-value understanding, unit-decade incompatibility is associated with increased response latencies and error rates (i.e., the so-called unit-decade compatibility effect) as for incompatible number pairs the separate comparisons of tens and units result in antidromic decision biases (e.g., Nuerk et al., 2001; Nuerk, Weger, & Willmes, 2004b; Korvorst &

Damian, 2008; Moeller, Fischer, Nuerk, & Willmes, 2009b; see Nuerk & Willmes, 2005 for a review; see also Nuerk et al., 2004b; Pixner, Moeller, Zuber, & Nuerk, 2009 for children data). Similar to the observed language-specific effects in transcoding, number word attributes like e.g., inversion have also been found to influence number magnitude comparison performance (see Nuerk et al., 2005 for adult data; Pixner, 2009, for children data).

Table 1: Place-value understanding and its consequences for the correctness of response for compatible and incompatible trials

		Unit-Decade Compatibility			
		Compatible		**Incompatible**	
		Example	**Response**	**Example**	**Response**
Place-value understanding	Correct (Decades most relevant – as indicated by capitalization)	57 V V 42	→ correct response as 5 > 4	62 V Λ 47	→ correct response as 6 > 4
	Incorrect (Units most relevant – as indicated by capitalization)	57 V V 42	→ correct response as 7 > 2	62 V Λ 47	→ incorrect response as 2 < 7

When place-value understanding is correctly applied decade digits are assigned a higher weight (indicated by capitalization) as their influence on the comparison outcome is theoretically ten times as large as the influence of the unit digits. Both compatible and incompatible items are correctly responded to when place-value understanding is correct (see upper row). However, in the case of incorrect place-value understanding (resulting in an over weighting of the units), the correctness of response differs (see lower row). For incompatible trials, such an incorrect *place-value understanding leads to an* incorrect *response: the smaller number with the larger units (e.g., 47) is wrongly categorized as larger because e.g., the unit 7 of 47 is larger than the unit 2 of 62. Importantly, this is different for compatible trials. A* correct *response might be observed even for* incorrect *place-value understanding because the units lead to the same response as the decades (e.g., 42_57; 4 < 5 and 2 < 7). So, even if the magnitude judgement is incorrectly based on the units, this can still yield a correct response in compatible trials. Therefore, such incorrect place-value understanding produces much more errors in incompatible trials. Consequently, the compatibility effect for errors is enlarged in children with a poorer, more often incorrect, place-value understanding.*

Please note that in contrast to Miura and colleagues (1994; 1989) we were not interested in the pure explicit knowledge of place-value principles as required to recode numbers into non-symbolic sets of tens and units blocks. Instead we paid particular attention to effects of the place-value structure of the symbolic Arabic number system inherent in specific numerical tasks. This means that we aimed at specifically focusing on the way place-value knowledge is applied in symbolic numerical tasks for which mastery of the place-value structure may be beneficial. Nevertheless, these tasks were not designed to explicitly assess children's knowledge of digital and positional values within a number by transformations into non-symbolic sets. Thus, the current study goes a step beyond the research by Miura and colleagues (1994; 1989) who showed that the strategies used to recode numbers into sets of tens and units blocks depend on language specificities in number word systems. However, Miura and co-workers did not investigate in which way these language effects as mediated by mastery of the place-value structure of the Arabic number system determine performance in other symbolic numerical tasks; and more particularly whether or not such differences implicitly influence the further development of arithmetical competencies such as addition.

In summary, recent research clearly indicates that mastery of the place-value structure of the Arabic number system is not only crucial for successful transcoding (see above) but is also essentially involved in tasks as easy as number comparison. In the final paragraph, the influences of basic place-value understanding on addition performance will be reviewed briefly.

The Arabic place-value structure and arithmetic performance

Comparable to the tasks introduced above the acquisition of more complex arithmetic problem solving (e.g., addition) might depend on a correct understanding of the place-value concept. Whenever the result of an addition problem contains more than one digit the single digits of the result have to be integrated into one coherent number according to the Arabic place-value structure. For instance, consider the problem 132 + 257. Here, unit, decade and hundred digits need to be manipulated and arranged correctly to obtain the result 389, i.e., overwriting of zeros: 300 and 89 → 389. Moreover, when addition problems become more difficult, e.g., by involving a carry operation, correct integration of digits into the place-value system gets more demanding as well. Considering the problem 132 + 259, this means that not only the correct result (i.e., 391) has to be integrated into the place-value structure. Additionally, one needs to keep track of the carry required at the unit position (i.e., $9 + 2 = 11$; $11 \geq 10$ → carry 1).

Thus, it is reasonable to assume that apart from the use of procedures, strategy selection and strategy efficiency (Imbo & Vandierendonck, 2008b) multi-digit addition involves a profound knowledge of basic numerical skills, in particular, the understanding of the place-value structure of the Arabic number system. Interestingly, Gervasconi and Sullivan (2007) reported that in first and second grade about 10% and 27% of children, respectively, exhibited problems in understanding the place-value structure of the Arabic number system. Hence, it is sensible to suppose that a better understanding of basic numerical concepts and especially of the place-value structure of the Arabic number system during early school years may have an influence on calculation performance later on[1].

Objectives and hypotheses

To test this prediction, the current study investigated whether third graders' addition performance as a measure of calculation ability was affected by their basic numerical knowledge, i.e., their place-value understanding as indexed by transcoding and magnitude comparison in first grade. It is generally agreed that calculation competencies are determined by both individual intelligence (e.g., Aiken, 1971; Hale, Fiorello, Bertin, & Sherman, 2003;

[1] Please note that we do not claim that the ability to name a number correctly to be a necessary prerequisite for calculation. However, we propose that the capability of naming a number correctly early on, involving mastery of the place-value structure of the Arabic number system, shall be a valid predictor of later arithmetic competencies such as addition performance.

Roid, Prifitera, & Weiss, 1993; see Sattler, 2001 for a review) as well as working memory capabilities (e.g., Geary, Hoard, Byrd-Craven, & DeSoto, 2004; McLean & Hitch, 1999; Swanson & Sachse-Lee, 2001; see Zago, Petit, & Turbelin, 2008 for adult imaging data). Moreover, influences of working memory capabilities have also been associated with children's development of numerical representations in recent models of typical as well as atypical development of numerical cognition (von Aster & Shalev, 2007; Rubinsten & Henik, 2009, respectively). Against this background it is important to account for rather unspecific effects of intelligence and working memory in our longitudinal analyses because we were only interested in the specific effects of early place-value understanding on later addition performance. So, by the inclusion of these variables (i.e., intelligence and working memory) we aimed at isolating these specific effects as the beta weights observed in the multiple regression analyses are then free from influences of intelligence and/or working memory and can be considered of reflecting the particular influence of children's early place-value understanding.

To particularly investigate influences of the Arabic place-value structure, this study focused on German speaking children. As outlined above, the German language inherits an inversion of tens and units in number words compared to digital notation. In the context of the current study this specificity may allow for a more comprehensive investigation of children's competencies in place-value understanding. As understanding the principle of reversed sequences of unit and decade-digits poses a particular complicacy (Zuber et al., 2009), a higher error variance for German-speaking participants can be expected compared to participants with other first languages (e.g., Brysbaert, Fias, & Noël, 1998). Since some precursors of mathematical achievement might remain undetected due to low-variance performances, the more not transparent German number-word system may thus allow for a more solid identification of basic numerical precursor competencies as it increases performance variance.

Following the progression of research approaches from one task – one representation (task approach) over the influence of specific numerical effects on overall performance to the investigation of the relevance of basic numerical effects for specific associated effects in later arithmetic (effect approach; see above) the hypotheses for the current study were three-staged:

In a (i) step it shall be investigated whether overall performance in basic numerical tasks in first grade (i.e., the overall error rate of magnitude comparison and transcoding) is a valid predictor of overall calculation performance in third grade. It is expected that children who commit few errors in magnitude comparison and/or transcoding also perform better in the calculation task (this analysis follows the task approach).

In the (ii) step the influence of specific numerical representations in first grade (i.e., place-value understanding) on overall addition performance in third grade shall be analysed (reflecting an effect approach of predictors; see also Holloway and Ansari, 2009, for a similar design). Specifically, it is expected that

a) difficulties in implementing the reversed order of tens and units in first grade, i.e., an increased number of pure inversion transcoding errors should be associated with a higher overall error rate in addition.
b) problems in unit-decade identification in first grade, i.e., a more pronounced unit-decade compatibility effect in number magnitude comparison is assumed to be accompanied by an increased overall error rate in addition.

In the (iii) step, analyses shall be focused on the question of whether specific numerical effects attributed to a single representation (i.e., pure inversion errors and place-value understanding) in first grade predict specific effects related to the same underlying representation in an arithmetic task two years later (i.e., pure inversion errors and the carry effect in addition). So far, only the predictors of numerical development have been operationalized in an effect based approach (e.g., Holloway & Ansari, 2009). Applying a similar approach to the criterion variables extends previous studies as to our knowledge the effect-based approach has never been employed in longitudinal studies on numerical development before. In particular, it is hypothesized that

c) the number of pure inversion transcoding errors in first grade should particularly predict difficulties in applying a required carry operation in addition, hence, specifically increasing the number of errors in addition problems involving a carry as well as the carry effect.
d) a relatively larger unit-decade compatibility effect indexing less elaborate identification of tens and units should also be associated with a specifically increased error rate in carry problems as well as a more pronounced carry effect.

METHOD

Participants: The current study was part of a large-scale project evaluating children's basic numerical development (see also Zuber et al., 2009; Pixner et al., 2009). In this longitudinal study, a total of 94 children (48 girls[2]) from five Austrian elementary schools were assessed on a variety of tasks (see below) by the end of grade 1 (mean age: 7 years 4 months, standard deviation (SD) = 7.1 months; range: 6 years 5 months to 8 years 7 months) and at the end of grade 3 (mean age: 10 years and 3 months, SD = 5.1 months; range = 9 years 4 months to 11 years 5 months). For selection purposes consent of school principles and school districts was obtained first. Then all parents of children in the respective schools were informed about the study and were kindly asked for the participation of their child. Only after parental informed consent was obtained the child was included in the study.

[2] Please note that the sample of first graders originally comprised 130 children. However, 36 children could not be considered for the assessment in third grade which represents a relatively high drop-out rate. As children were recruited from five elementary schools different reasons may account for this: children either had to repeat a grade, did no longer attend the respective school or no signed consent form was provided. Nevertheless, the drop-out did not bias the results in first grade systematically as both the distribution of transcoding errors as well as the compatibility effect did not differ substantially between the original and the reduced sample of first graders. At both occasions children were assessed individually during school hours in one-on-one sessions in a quiet room. All tasks were administered in German, the native language of all participating children.

As the current manuscript focuses on addition as an indicator of arithmetic abilities the Austrian curriculum concerning addition and its counterpart subtraction, which are introduced simultaneously, shall be described briefly. In Austria children should master numbers up to 20 as well as additions and subtractions within this range by the end of grade one. At the beginning of grade two the to-be-mastered number range is extended to 100 and after a recap of single-digit addition and subtraction mental addition and subtraction of two-digit numbers is introduced. Finally, in grade three the written procedures for two-digit addition and subtraction are dealt with.

Tasks, stimuli and procedure:

First grade assessment: In first grade, children were administered a transcoding task as well as a number magnitude comparison task, amongst others; lasting about 10 to 15 minutes each. Additionally, general cognitive measures such as intelligence (CFT-1, Cattell, Weiß, & Osterland, 1997) and WM capacity (paradigms to be described below) were assessed. For all numerical tasks employed in the current study Spearman-Brown corrected split-half reliability was computed in an item analysis on error rates.

(i) In the *Transcoding task* children were asked to write down 64 numbers to dictation. The item set consisted of 4 one-digit numbers (e.g., 4), 20 two-digit numbers (e.g., 15, 78), and 40 three-digit numbers (e.g., 281, 306; for a detailed description of the stimuli please refer to Zuber et al., 2009). Errors were coded according to the five disjoint subgroups of transcoding errors as introduced by Zuber et al. (2009): (i) lexical errors (substitution of lexical primitives, e.g., 71 → 81), syntactic errors such as (ii) pure inversion errors (e.g., 71 → 17), (iii) additive composition errors (e.g., 171 → 10071), (iv) multiplicative composition errors (e.g., 571 → 5171) and (v) combination errors (e.g., combination of additive composition and inversion: 571 → 50017). In its current version a split-half reliability of r_{tt} = .95 was obtained for the transcoding task.

(ii) In the *magnitude comparison task* children had to evaluate 120 two-digit number pairs and indicate which one of the two numbers presented above each other on a computer screen was larger by pressing the corresponding response key. The two to-be-compared numbers were presented in Arial font (size: 60) until one of the two response buttons was pressed. The stimulus set comprised 80 between-decade trials for which decade distance (i.e., small vs. large; e.g., 23_47 vs. 21_85) and unit-decade compatibility (i.e., compatible vs. incompatible; e.g., 42_57 vs. 47_62) was manipulated in a 2 x 2 within participant design. Additionally, the stimulus set involved 40 within-decade trials (e.g., 32_37). The split-half reliability of the magnitude comparison task was r_{tt} =.45.

(iii) The *WM tasks* were designed to tap the three WM components suggested by Baddeley (1986): verbal, visuo-spatial WM and central executive: a) *verbal WM* was assessed by a letter repetition task. This task involves the presentation of spoken sequences of letters for immediate serial recall. After a practice trial, a maximum of three lists were presented to the children at each length, starting with a two-item sequence. Only when two of three lists of a particular length were recalled correctly, list length was increased by one otherwise testing was stopped. Letters were presented at a rate of approximately one per second. The maximum sequence length at which two lists were correctly recalled was then used a measure of verbal WM span. b) *Visuo-spatial WM* was measured by the Corsi block tapping task (Corsi, 1972). The location of a certain number of cubes on a board had to be remembered and reproduced in correct order. After a practice trial subsequent procedure and classification of results was similar to that for the letter repetition task. c) For *central executive (CE) assessment* each child had to recall sequences of spoken letters and Corsi blocks in reversed order (for a similar consideration of backward spans as a central executive task see e.g., Gathercole & Pickering, 2000). Practice trials ensured that the concept of “reverse” was correctly understood by each child. Subsequent procedure and scoring were identical to that described above for forward span tasks.

Follow-up assessment: Two years later (i.e., in third grade) children's calculation ability was assessed by an *addition task*. The addition task comprised 48 two-digit addition problems of which 24 required a carry operation, lasting approximately 10 to 15 minutes. In the applied choice reaction version children had to indicate which one of two proposed solution probes was correct. Incorrect probes deviated from the correct result by either ±2 or ±10 to prevent parity based solution strategies. Additionally, average split of incorrect solution probes was zero for both carry and no-carry problems and problem size was matched between the item categories. Problems were presented in the form xx + xx at the x-/y-coordinates (0/-120) while the two probes appeared at (-250/120) and (250/120) with screen resolution set to 1024 x 768. Both, addition problems as well as solution probes were presented simultaneously in white against a black background (font: Arial; size: 60) until one of the two corresponding response buttons was pressed: when the left probe was correct the *Alt* button of a standard keyboard had to be pressed whereas the *Alt Gr* button had to be pressed when the right probe was correct. No feedback was given as to the correctness of children's responses. After one of the response buttons had been pressed and the next item was presented following an inter-stimulus interval was 560 ms. Prior to the 48 critical trials each child completed 10 practice trials to familiarize with display layout and task requirements. For the addition task a split-half reliability of $r_{tt} = .91$ was observed.

Analysis: Performance in all numerical tasks incorporated in the analyses was measured in terms of error rates. As error rates cannot be assumed to be normally distributed error rates were arcsine transformed [i.e., 2*arcsine(sqrt(error rate))] to approximate normal distribution. Error rates instead of reaction times were chosen due to two main arguments: First, reaction times obtained from children often show extreme intra- as well as inter-subject variance that could drive effects in one or another direction. Against the background of the problem of capitalizing on differences in and between individual variance by using reaction times in children, error rates seemed to the more reliable measure of task performance to us.

Second, we did not record reaction times for the transcoding task as it is questionable up to which point in time reaction time should be defined. Considering response onset (i.e., starting to write down the first digit) seems inappropriate as transcoding, especially for multi-digit numbers, is a lengthy and often sequential process in which the correct digital symbols have to be arranged in the correct order. Thus, such a measure of reaction time would not reflect the end of the transcoding process but only a sub-step on the way to the correct result. On the other hand, taking the time until the answer is completed would be to measurement error due to differences in motor and writing skills, the child being distracted or even restarting transcoding after losing track of the solution. Finally, in most cases only reaction times of correctly solved trials are analyzed. However, in the current study we were explicitly interested in what kind of errors the children committed. Most importantly, reaction times are not capable of coding what kind of error a child made, which was the critical question in the current study. Therefore, it was decided to run all subsequently described analyses on arcsine transformed error rates.

Please note that incorporating gender into the analyses revealed no reliable influences of gender, neither in the factorial nor in the regression analyses. Therefore, the data for boys and girls was pooled for all analyses reported.

RESULTS

Intelligence and Working Memory

In first grade, average intelligence level as assessed by the CFT 1 (Cattell et al., 1997) was T = 59 (95% CI: 57.7 – 60.9). At the same time children's average score in the visuo-spatial WM task (Corsi block) was 4.10 (95% CI: 3.96 – 4.23). In the verbal WM task (letter repetition) the average score was 4.31 (95% CI: 4.17 – 4-44). The mean score of both backward span tasks was chosen as an index for the CE (see Barrouillet, Mignon, & Thevenot, 2008 for a similar approach). Hence, the mean CE score was 2.70 (95% CI: 2.57 – 2.84).

Numerical task results

The results of the transcoding task and the magnitude comparison task administered in first grade were reported in great detail and discussed extensively in Zuber et al. (2009) and Pixner et al. (2009), respectively. However, this study focuses on the longitudinal relevance of first-grade performance. Therefore, the first-grade results will only be briefly summarized below to provide the reader with the information necessary to follow the subsequent longitudinal analyses.

Transcoding in first grade

Generally, the error pattern in the longitudinal sample of the current study (n = 94) was similar to that reported by Zuber et al. (2009; n = 130). For the reduced sample Bonferroni-Holm corrected *t*-tests (Holm, 1979) showed that combination errors and additive composition errors were the most frequent error types [M = 16.7% errors (95% CI: 13.2% – 20.2%) vs. M = 15.5% errors (95% CI: 12.2% – 18.8%), respectively; $t_{\text{combination vs. add. composition}}(93) = 0.93$, $p = .35$, $d = .07$; all other $t > 2.84$, all $p < .01$, all other d > .62] followed by pure inversion errors (M = 7.1%; 95% CI: 5.7% – 8.6%), lexical errors (M = 2.7%; 95% CI: 2.0% – 3.4%) and multiplicative composition errors (M = 0.8%; 95% CI: 0.2% – 1.4%, see Figure 1).

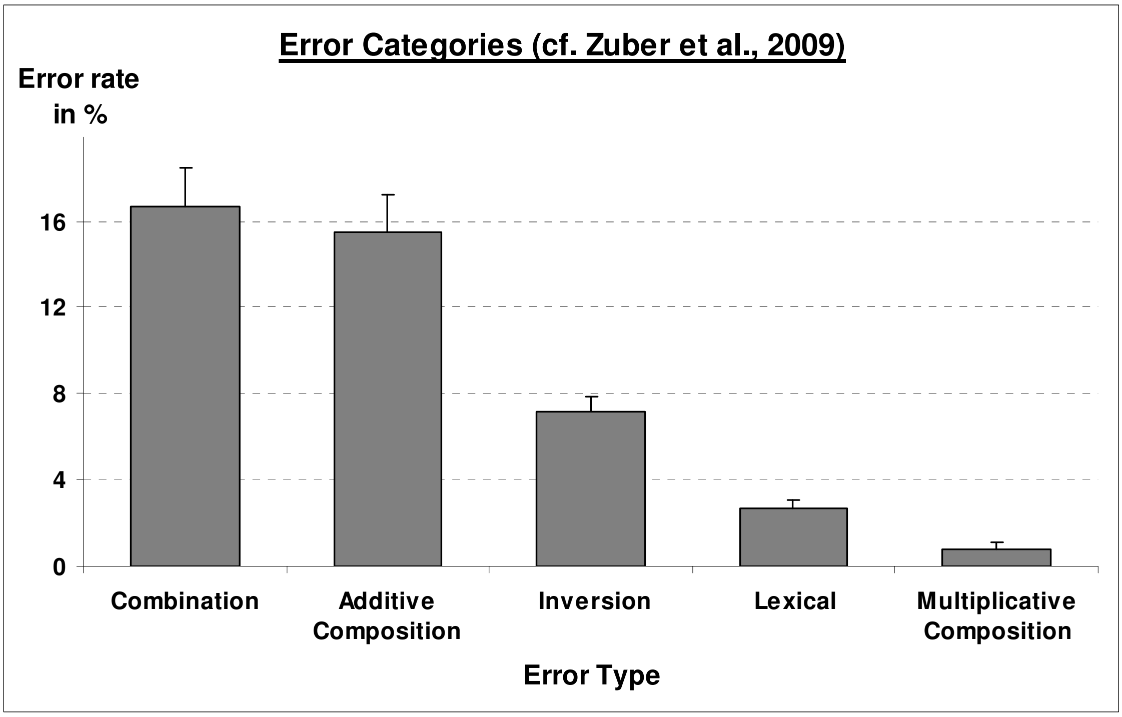


*Figure 1: Distribution of error types as categorized by Zuber et al. (2009) in the reduced sample (*n *= 94) of the current study. Please note that the error rates given reflect the respective percent of erroneous items in reference to all transcoding items and thus are not supposed to add up to 100%.*

Magnitude comparison in first grade

An ANOVA showed that both the compatibility effect as well as the (decade) distance effect were highly significant for errors in the reduced sample [distance: $F(1, 93) = 36.97$, $p < .001$, $\eta_p^2 = .28$; compatibility: $F(1, 93) = 88.61$, $p < .001$, $\eta_p^2 = .49$]. This indicated that first graders' classifications in the magnitude comparison task were more accurate for number pairs involving a large compared to a small (decade) distance [large: M = 18.1% errors (95% CI: 14.6% – 21.6%) vs. small: M = 23.2% errors (95% CI: 19.6% – 26.8%)] as well as for compatible than for incompatible number pairs [compatible: M = 9.6% (95% CI: 7.2% – 12.0%) vs. incompatible: M = 31.6% errors (95% CI: 26.1 – 37.2%].

Please note that even for the most difficult so-called incompatible number pairs children clearly performed above chance level. As in these trials the larger number contains the smaller unit identification of tens and units and their associated value is inevitably necessary to solve the task successfully. Almost 70% correct responses indicate, that the participating children understood the place-value principle at least implicitly, i.e. that the decade digits have a higher importance than the unit digits. Thus, these children were not assessed too early in their numerical development.

Addition task in third grade

The influence of a required carry procedure on error rates in the addition task was appraised using a paired samples *t*-test. The *t*-test revealed that children committed reliably more errors in those addition problems which required a carry operation (M = 11.7% errors; 95% CI: 9.7% - 13.6%) compared to problems not requiring a carry [M = 9.7% errors; 95% CI: 7.6% - 11.8%; $t(93) = 2.15$, $p < .05$, $d = .24$].

Longitudinal analysis: Influence of basic numerical competencies on calculation ability

(i) Task approach analysis: Interrelations of overall performance of basic numerical task and mental arithmetic

In this first step, we were interested whether addition performance in general (in terms of error rate) was determined by general transcoding performance and/or overall performance in the number magnitude comparison task. This analysis followed the task approach (see above), as overall error rates of the respective tasks serving as measures of task performance were Pearson correlated to evaluate a possible interrelation between magnitude comparison and transcoding as well as the predictive power of the latter two tasks on overall addition task performance in grade three. Additionally, the differential influences of these two variables were examined in a multiple regression analysis. Apart from above described variables measures of intelligence, verbal as well as visuo-spatial WM and CE were considered in correlation as well as regression analyses of each of the two steps.

The correlation analysis of overall performance (in terms of error rate) showed that performance in transcoding as well as in number comparison as assessed in first grade was positively correlated with overall performance in the third grade addition task. This indicated that children who committed relatively more errors in either transcoding or number comparison also exhibited a higher error rate in an addition task administered two years later (see Table 2, Panel A). Furthermore, in the current data set no reliable association between addition performance and either intelligence or WM capacity could be observed (see Table 2, Panel A). So, only the numerical predictors in first grade were associated reliably with addition performance in third grade, while the non-numerical variables did not predict any variance in the addition task.

When running a multiple stepwise regression analysis incorporation all variables of the correlation matrix in Table 1, Panel A to evaluate the differential influences of overall comparison and/or transcoding performance on overall addition performance, overall performance in the magnitude comparison task turned out to be the only reliable predictor [$b = .25$, $t = 2.48$, $p < .05$] incorporated in the model [$R = .25$, adjusted $R^2 = .05$, $F(2, 92) = 6.13$, $p < .05$, $f^2 = .07$]. As already indicated by the positive correlation, a higher error rate in the magnitude comparison task in grade one was associated with a higher error rate in the addition task two years later. The fact that overall transcoding performance was not found to be a reliable predictor of overall addition performance may be attributed to the very high intercorrelation of overall performance in magnitude comparison and transcoding (see Table 2, Panel A).

.

Table 2: Correlation matrix depicting the influence of basic numerical tasks (Panel A) as well as the influence of specific numerical effects (Panel B) on overall addition performance

A	Task	1.	2.	3.	4.	5.	6.	7.
Numeric	1. Addition (3rd grade)	-	.25*	.24*	.01	.03	.09	- .12
	2. Comparison (1st grade)		-	.54**	- .08	.00	- .20(*)	- .33**
	3. Transcoding (1st grade)			-	- .19(*)	.00	- .27*	- .33**
Non-numeric	4. Intelligence (1st grade)				-	.00	- .04	.21*
	5. Verbal WM (1st grade)					-	.27**	.14
	6. Spatial WM (1st grade)						-	.38**
	7. CE (1st grade)							-

B	Task/Effect	1.	2.	3.	4.	5.	6.	7.	8.
Numeric	1. Addition (3rd grade)	-	- .17(*)	.24*	.35**	.00	.03	.09	- .12
	2. Distance effect (1st grade)		-	- .04	.05	- .19(*)	- .19(*)	.03	- .07
	3. Compatibility effect (1st grade)			-	.13	.04	.04	- .01	- .21*
	4. Inversion (1st grade)				-	.01	- .19(*)	- .16	- .25*
Non-numeric	5. Intelligence (1st grade)					-	.00	- .04	.21*
	6. Verbal WM (1st grade)						-	.27**	.14
	7. Spatial WM (1st grade)							-	.38**
	8. CE (1st grade)								-

(*) = $p < .10$; * = $p < .05$; ** = $p < .01$

(ii) Effect approach analyses of predictors: Predictions from specific basic numerical competencies on overall performance in mental arithmetic

In the second more differential analysis the relevance of specific effects observed in first grade assessment and supposed to reflect specific mastery of an associated numerical representation for overall addition performance was investigated (i.e., an effect approach for the numerical predictors; see Holloway & Ansari, 2009 for a similar approach). In particular, the influence of the (decade) distance effect (reflecting number magnitude understanding), the compatibility effect (reflecting integration of tens and units) as well as the percentage of pure inversion errors in transcoding (reflecting place-value understanding) on the overall error rate in the addition task were evaluated in correlation and multiple regression analyses Again, measures of intelligence, verbal as well as visuo-spatial WM and CE were considered in correlation as well as regression analyses.

Following the approach suggested by Holloway and Ansari (2009) the influence of specific numerical effects indexing specific basic numerical competencies on addition performance were investigated in a correlation analysis. In contrast to Holloway and Ansari (2009) we observed the distance effect (reflecting number magnitude understanding) to be negatively correlated with addition performance in regard of error rate (see Table 2, Panel B). So, children exhibiting a comparably larger (decade) distance effect in grade one also committed less errors in the addition task in grade three. In addition, we also identified indices of early place-value understanding in first grade (such as the compatibility effect and the percentage of pure inversion errors) to be correlated with addition performance two years later (see Table 2, Panel B). This meant that children with a more pronounced compatibility effect in magnitude comparison also had a relatively higher error rate in the third grade addition task compared to children with a small compatibility effect. Furthermore, those children who committed comparably more pure inversion errors in the transcoding task also exhibited a higher error rate in third grade addition.

Again, a multiple stepwise regression analysis was conducted to evaluate the differential influences of the variables depicted in Table 2, Panel B on overall addition performance in third grade. The final model [R = .44, adjusted R^2 = .16, $F(4, 90) = 6.99$, $p <$ .001, f^2 = .23] included the three predictors a) number of pure inversion transcoding errors, b) compatibility effect, and c) (decade) distance effect (see Table 3). Inspection of the beta weights indicated that children who committed more pure inversion transcoding errors in first grade also exhibited a higher overall error rate in the third grade addition task. Similarly, children with a relatively large compatibility effect in the first grade magnitude comparison task committed more errors in the addition task two years later. This indicated that in both cases flawed mastery of the place-value concept in first grade was associated with an increased overall error rate in the addition task in third grade. Finally, in contrast to the findings of Holloway and Ansari (2009) a larger (decade) distance effect in first grade implied better overall performance in the third grade arithmetic task.

Table 3: Regression model predicting the overall number of errors in the addition task on the basis of specific numerical effects

Predictor	B	95% CI	*b*	Change in R^2	*t*	sig.	raw	partial
Constant	.45	.34 – .56			8.19	< .001		
Inversion errors (1st grade)	.30	- .13 – .48	.33	.12	3.47	< .001	.35	.34
Compatibility effect (1st grade)	.09	- .003 – .18	.18	.04	1.93	< .05$^{(*)}$	.24	.20
Distance effect (1st grade)	- .18	- .38 – .01	- .18	.03	1.87	< .05$^{(*)}$	- .17	- .19

CI = Confidence interval; $^{(*)}$ = tested one-sided

(iii) Effect approach analyses of predictors and criterion variables: Influence of specific basic numerical competencies on related effects in mental arithmetic

Finally, taking a step beyond Holloway and Ansari (2009) the effect approach was generalized to both, predictor and criterion variables. That is, we examined in how far mastery of specific basic numerical competencies, in particular place-value understanding, serves as a predictor for specific effects in later arithmetic performance. In particular, it was investigated whether the carry effect for addition (as requiring correct place-value integration) in grade three can be traced back to indices of basic numerical competency in first grade, such as the distance effect and, more particularly, to indices of place-value understanding such as the compatibility effect or the rate of inversion errors. Additionally, we were interested whether the percentage of errors in carry problems is accounted for best by the same basic numerical competencies as is the error rate in non-carry problems. To evaluate these most differential aspects of longitudinal numerical development three stepwise regression analyses were conducted. All regression analyses incorporated the same predictors but varied with regard to the dependent variable (i.e., carry effect for addition errors vs. error rate in carry problems vs. error rate in non-carry problems). The included predictors were: (i) the (decade) distance as well as (ii) the compatibility effect for errors in the magnitude comparison task, the five disjoint subgroups of transcoding errors as introduced by Zuber et al. (2009): (iii) lexical errors, (iv) inversion errors (v) additive composition errors, (vi) multiplicative composition errors and (vii) combination errors. In addition to these predictors reflecting basic numerical competencies number-unrelated measures of IQ, verbal as well as visuo-spatial WM and CE were also incorporated as predictors. Thereby, it was aimed to distinguish possible non-numerical influences of intelligence or WM capacity from effects attributable to children's basic numerical competencies such as transcoding, number magnitude as well as place-value understanding.

Please note that for this last step only the results of the regression analyses will be reported in the main text body for reasons of lucidity and brevity. To provide the reader with a full overview over all results nevertheless, all correlation analyses are given in Appendix A.

Carry effect: The stepwise multiple regression analysis produced a significant model [R = .29, adjusted R^2 = .06; $F(3, 91) = 4.08$, $p < .05$, $f^2 = .09$] incorporating the first grade CE score and the percentage of pure inversion transcoding errors as the only reliable predictors (see Table 4, Panel A). The positive beta weight for both predictors indicated that a relatively higher score in the CE working memory task as well as a higher percentage of pure inversion errors in first grade was associated with a more pronounced carry effect in third grade. Neither the (decade) distance effect nor the compatibility effect was found to predict the size of the carry effect reliably (see Table 4, Panel A). However, contrary to what one would assume intelligence [all t = 0.36, all p = .72] was not a reliable predictor for the size of the carry effect.

To further investigate the hypothesized specific influence of the ability to successfully integrate tens and units into the place-value structure of the Arabic number system on carry problems compared to non-carry problems the regression analysis was rerun on the percentage of errors separately for carry as well as non-carry problems.

Carry problems: For the number of errors for the carry problems the regression analysis identified a model [R = .46, adjusted R^2 = .19; $F(4, 90)$ = 8.14, p < .001, f^2 = .27] incorporating the predictors a) percentage of pure inversion errors, b) visuo-spatial WM span, and c) compatibility effect, each as observed in first grade (see Table 4, Panel B). In line with our hypotheses, children who committed relatively more inversion errors in first grade also exhibited a higher error rate for carry problems in third grade. Additionally, responses of children with a relatively more pronounced compatibility effect for errors in grade one were also observed to be more error prone for carry problems (see Table 4, Panel B). However, again, a WM measure was positively correlated with the dependent variable: Children who scored higher on the visuo-spatial WM task in grade one also made comparatively more errors in third grade addition problems involving a carry operation. Comparable to the results for the carry effect the (decade) distance effect was not a reliable predictor of error rate in carry problems (see Table 4, Panel B). Again, intelligence was not a reliable predictor in the regression model [all t = 0.79, p > .43].

Non-carry problems: For the percentage of errors in the non-carry problems the results were completely different. The final regression model comprised the predictors percentage of combination errors and (decade) distance effect in first grade [R = .36, adjusted R^2 = .11; F(3, 91) = 9.81, p < .01, f^2 = .15]. A closer look at the beta weights showed that the percentage of combination errors in first grade was positively associated with the percentage of errors in no carry addition problems in third grade (see Table 4, Panel C). So, children who committed relatively more combination error in the transcoding task administered in grade one also had a higher error rate for no-carry addition problems in third grade. Additionally, children exhibiting a more pronounced (decade) distance effect in grade one committed fewer errors in non-carry addition problems in third grade. However, for both effects reflecting specific place-value understanding, i.e., the rate of inversion errors as well as the compatibility effect, no significant influence on the error rate for non-carry problems could be observed (see Table 4, Panel C). As for above reported analyses, intelligence was not considered as a reliable predictor in the regression model for non-carry problems (all t < 0.07, p > .95).

Table 4: Regression model predicting the carry effect in third grade (Panel A), the number of errors for carry problems (Panel B) as well as the number of errors in non-carry problems (Panel C) on the basis of specific numerical effects

A	Predictor	*B*	95% CI	*b*	Change in R^2	*t*	sig.	raw	partial
Constant		- .43	- .81 – - .09			2.27	< .05		
CE (1st grade)		.15	.03 – .27	.26	.04	2.47	< .05	.20	.25
Inversion errors (1st grade)		.25	.003 – .49	.21	.04	2.00	< .05	.14	.21
Distance effect (1st grade)		-	-	.06		0.56	.58	.05	.06
Compatibility effect (1st grade)		-	-	.02		0.14	.89	- .01	.02
B	Predictor	*B*	95% CI	*b*	Change in R^2	*t*	sig.	raw	partial
Constant		- . 06	- .49 – .38			.26	.80		
Inversion errors (1st grade)		.42	.21 – .62	.39	.14	4.03	< .001	.38	.39
Visuo-spatial WM (1st grade)		.10	.01 – .20	.20	.04	2.11	< .05	.14	.22
Compatibility effect (1st grade)		.10	- .004 – .20	.18	.03	1.91	< .05(*)	.23	.20
Distance effect (1st grade)		-	-	- .15	-	1.60	.12	- .13	- .17
C	Predictor	*B*	95% CI	*b*	Change in R^2	*t*	sig.	raw	partial
Constant		.43	.30 – .55			6.76	< .001		
Combination errors (1st grade)		.22	.08 – .35	.31	.10	3.20	< .01	.31	.32
Distance effect (1st grade)		- .23	- .49 – .02	- .18	.13	1.82	< .05(*)	- .17	- .19
Inversion errors (1st grade)		-	-	.10	-	0.94	.35	.21	.10
Compatibility effect (1st grade)		-	-	.11	-	1.07	.29	.23	.11

CI = Confidence interval; (*) = tested one-sided

DISCUSSION

The aim of the current study was to investigate the influence of early basic numerical competencies such as transcoding or number comparison on later arithmetic capabilities in typically developing children. Therein, we set out to resolve several limitations in the hitherto investigation of this developmental trajectory: (i) most of the existing evidence supporting this notion comes from studies employing a simpler task approach in which one particular representation is indexed by an overall performance measure (e.g., error rates) of one particular task. (ii) The stimulus sets used to assess basic numerical competencies in previous studies mostly comprised one-digit numbers only instead of multi-digit numbers (e.g., Holloway & Ansari, 2009). (iii) Consequently, when only single-digit stimuli were used examination of the influence of place-value understanding as an important precursor of later arithmetic ability was not possible in past research. (iv) Finally, a developmental trajectory from basic numerical concepts to arithmetic has never been observed in a longitudinal approach to date.

In addressing these issues our proceeding was three-staged: (i) We were interested whether overall performance in number comparison and/or transcoding as assessed in first grade were reliable predictors of overall addition performance in third grade. (ii) More specifically, attention was paid to whether mastery of specific basic numerical concepts in first grade (e.g., place-value understanding) as indexed by related numerical effects (e.g., compatibility effect) predicts overall addition performance two years later (see Holloway & Ansari, 2009, for a similar approach). (iii) Most specifically, the quality of such specific basic numerical effects (e.g., compatibility effect) serving as predictors for specific effects in later addition (e.g., carry effect) that can be attributed to the same underlying concept of place-value understanding, was of interest.

Our results were straightforward on each of these questions:

(i) We clearly identified overall transcoding and number comparison performance in first grade to be precursor competencies of general arithmetic performance in third grade.

(ii) When following the approach suggested by Holloway and Ansari (2009) we were not able to replicate their findings concerning the (decade) distance effect but took a successful next step in identifying precursors of arithmetic capability. On the one hand, we observed a more pronounced (decade) distance effect in first grade to be associated with a lower error rate in third grade arithmetic, whereas Holloway and Ansari (2009) found that a relatively smaller distance effect was associated with better mathematics performance (see below for detailed interpretation of these differences). On the other hand, we extended the results of Holloway and Ansari (2009) in that reliable influences of both the compatibility effect as well as of the number of pure inversion errors in first grade on overall arithmetic performance were present in our data: a larger compatibility effect and more pure inversion errors in transcoding implied a higher error rate in third grade arithmetic performance.

Thereby, the present data allowed for particularly pinpointing measures reflecting understanding of the place-value structure of Arabic numbers for both tasks to be most relevant for the prediction of overall arithmetic performance in later numerical development.

(iii) Finally, examining the influence of specific numerical effects in first grade (e.g., compatibility effect, pure inversion errors) on corresponding effects in third grade arithmetic (e.g., carry effect, error rate in carry problems) revealed several meaningful interrelations.

The size of the *carry effect* in third grade addition was predicted reliably by the first grade measure of CE functioning and the number of pure inversion transcoding errors: A higher number of pure inversion errors in first grade transcoding reliably predicted a more pronounced carry effect in third grade addition. Thus, flawed understanding of the place-value positions of tens and units in first grade is related to difficulties in later arithmetic performance in additions requiring a carry operation in which correct integration of tens and units is essential. We suggest that both variables tap the same underlying concept of place-value understanding in a basic numerical task in first grade (i.e., pure inversion errors in transcoding) as well as in a complex numerical task in third grade (i.e., addition involving carry over). Thereby, the present data indicate that if the place-value concept is not mastered in early schooling difficulties in the successful application of this concept will evolve in later numerical development. To our knowledge, these are the first longitudinal data which directly evidence the importance of early place-value understanding. Additionally, the CE component of WM was also positively correlated with the size of the carry effect: the better the CE in first grade the larger the carry effect in third grade indicating an influence of WM on complex addition which has previously also been observed in adults (e.g., Imbo et al., 2007a; 2007c).

The regression model predicting the number of errors committed in *carry addition problems* incorporated three reliable predictors: a) the number of pure inversion errors in transcoding, b) visuo-spatial WM span and c) the size of the compatibility effect in number comparison. So, children who had problems to identify tens and units successfully in transcoding and/or who showed relatively larger interference effects due to a worse integration of tens and units in two-digit number comparison also committed particularly more errors on addition problems requiring a carry operation. Again, poor performance in a task specifically involving place-value updating and integration is predicted best by basic numerical competencies reflecting exactly these requirements. Moreover, a higher visuo-spatial WM span in first grade was associated with an increased number of errors for carry problems. This positive correlation is not in line with standard hypotheses as one would assume a larger WM span to be associated with better performance. We will elaborate on this finding below.

Finally, for the number of errors for *non-carry addition problems* the number of combination errors in transcoding and the size of the (decade) distance effect turned out to be the only reliable predictors: In line with the results of the regression analysis for overall addition performance a larger distance effect in first grade was associated with fewer errors for non-carry addition problems indicating a major influence of the representation of number magnitude on easier problems. Furthermore, children with a higher number of combination errors also committed more errors in non-carry addition problems. Importantly, the fact that combination errors and not pure inversion errors were the best predictor of performance for non-carry problems may indicate that for these simpler problems demands on place-value understanding are not as specific as for carry problems. Obviously, place-value understanding is required for all two-digit addition problems. However, incorrect ordering of tens and units, as reflected by pure inversion errors, is particularly detrimental for carry problems: To correctly solve a carry problem the units must be summed first. Subsequently, the tens and the carry from the units must be summed to obtain the correct result. Contrarily, the order of computations is irrelevant for non-carry problems as there is no carry from the units to the tens. So, whether the tens or the units are summed first yields identical results as long as the position of the single digits within the place-value system is applied correctly. For this reason, it is conclusive that the number of inversion errors (indexing wrong order in transcoding) specifically predicted performance for carry problems while the less pure category of combination errors became a predictor for non-carry problems. Such an interpretation corroborates our argument that understanding the place-value structure of Arabic numbers should be of particular relevance in situations requiring place-value updating and/or integration.

Taken together, the results of the present longitudinal evaluation clearly indicated early place-value understanding to be a highly predictive precursor competency of later arithmetic performance in general and in particular for the capability to integrate tens and units into the Arabic place-value system as required in basic arithmetic operations such as addition problems requiring a carry. Yet, at the same time intelligence was not considered to be a reliable predictor of task performance in any of our analyses. At a first look this might seem awkward. However, this observation is in line with recent data that also showed no reliable interrelation between intelligence and basic numerical tasks such as number magnitude comparison (De Smedt et al., 2009b; Halberda, Mazzocco, & Feigenson, 2008) as well as number reading (De Smedt et al., 2009b) and transcoding (Zuber et al., 2009). Most importantly, this finding seems invariant to the fact which intelligence scale was used. While Zuber and colleagues (2009) also used the CFT-1 and De Smedt and co-workers (2009b) employed a comparable matrices test (i.e., Raven's standard progressive matrices, Raven, Court, & Raven, 1992), Halberda et al. (2008) measured intelligence by the WASI (Wechsler abbreviated scale of intelligence, Wechsler, 1999). Synced with the fact that there were no peculiarities during the testing sessions at any of the involved schools, the current data do not corroborate the notion of a direct link between intelligence and basic numerical concepts such as place-value understanding, thereby replicating previous results (De Smedt et al., 2009b, Halberda et al., 2008).

A word on the validity of the current study

Although statistically reliable influences of place-value understanding in first grade on arithmetic performance in third grade were observed in the current study, the results tell little about the external validity of these influences. To evaluate the validity of the current results we ran an additional regression analysis involving third grade mathematics school grades as the dependent variable. Unfortunately, it was not possible to get the mathematics grades from all children due to reasons of data protection. Nevertheless, we were able to collect the mathematics grades of a subsample of 54 children. A stepwise regression analysis was conducted predicting the mathematics grade by the same predictors as incorporated in all other analyses with one exception. Due to the reduced sample we only differentiated between inversion related and inversion unrelated transcoding errors. The group of inversion related errors comprised all errors (including combination errors) that involved an error in digit order in at least one position, while all other errors were pooled in the category inversion unrelated errors. In line with our argument that place-value understanding may be a crucial building block for later arithmetic competencies it was observed that inversion related transcoding errors assessed in first grades were the only reliable predictor of mathematics grade two years later (R = .50, adjusted R^2 = .22; $F(2, 51)$ = 8.59, p < .001, f^2 = .33; b = .40, t = 3.15, p < .01) whereas inversion unrelated errors did not account for any additional variance (b = .15, t = 1.13, p = .27). Based on this result we are confident that our tasks provide a reliable (see above) and valid estimate of children's numerical competencies.

The fact that the compatibility effect was not found to be a reliable predictor of mathematics grade (b = - .05, t = 0.38, p = .71) might be caused by the lower reliability of the magnitude comparison task as compared to the transcoding task. The reliability of the magnitude comparison task might be lower because it involved numbers up to 100 and was thus rather difficult in general. In this context, it has to be noted that by the end of grade one (i.e., the time of assessing our participants) children in Austria are supposed to just master the numbers up to 20 and their arithmetic relations. Nevertheless, in the task approach analysis the magnitude comparison task was a reliable predictor of later addition performance. This seems to imply that this significant predictive power we observed could eventually be augmented and maybe even specified to the compatibility effect when the magnitude comparison task is made more reliable (e.g., by employing more and different stimuli).

The distance effect and later arithmetic performance

Recent results concerning the influence of the distance effect are controversial. On the one hand, Holloway and Ansari (2009) observed a larger distance effect to be associated with worse performance in typically developing children (see also De Smedt et al., 2009b for similar results; but see Nuerk et al., 2004a for contrasting findings). Additionally, Delazer and colleagues (2006) reported a larger distance effect in a neurological patient (suffering from posterior cortical atrophy) exhibiting difficulties in magnitude representation and arithmetic. Finally, Kaufmann and Nuerk (2008) found that children with ADHD-C also had a larger distance effect than their non-ADHD peers, which has been interpreted as reflecting group differences in number representations (i.e., a larger distance effect possibly being due to more noisy number magnitude representations).

However, in the present study, a positive correlation between the distance effect and performance was obtained suggesting a larger distance effect to be associated with better performance. Nevertheless, such a (cor)relation is corroborated by the results of Rousselle and Noël (2007). Comparable to the present data these authors also report a relatively larger distance effect being associated with better arithmetic performance. As Rousselle and Noël (2007) investigated a sample of dyscalculic children they attributed this finding the possibility of some children relying on some "peculiar strategies" (Rousselle & Noël, 2007, p. 387) when comparing two numbers. Such back-up strategies (e.g., counting) may have been necessary as direct access to the numbers' magnitude representation was impaired. This interpretation can be adapted for the results of the present study (as well as for results reported by Nuerk et al., 2004a). The number comparison task administered in grade one required children to single out the larger of two two-digit numbers ranging from 12 to 98. As by the end of grade one only the numbers up to 20 and their interrelations were taught at schools in Austria this opens the possibility that some children may have employed strategies other than magnitude comparisons (e.g., counting). Hence, when eliminating all participants with a negative distance effect to account for this possibility (see Rousselle & Noël, 2007 for a similar proceeding) the (decade) distance effect was no longer a reliable predictor of either overall addition performance or error rate for non-carry problems (both $t < 0.95$, both $p > .35$). Thus, the unexpected influence of the (decade) distance effect on arithmetic performance may be associated with the difficulty of the number comparison task (for a similar argumentation see Nuerk et al., 2004a).

In fact, in our view, these results suggest that the relationship between the distance effect and arithmetic performance may not be monotone (see Figure 2). What are the prerequisites to obtain such a monotone relation? One needs a large magnitude effect to explain much variance of numerical distance *and* one needs low error variance attributable to the fact that all participants need to process numerical magnitude in a similar analogue way, without using any back-up strategies producing much error variance and thereby making the effect more fuzzy. We assume that worse access to analogue magnitude is associated with a larger distance effect (see Kaufmann & Nuerk, 2008, for an elaboration of that argument). However, we also assume that when access to magnitude is not always possible anymore in an analogue form this may result in much larger error variance (e.g., because of back-up counting strategies). This increased error variance may then account for a reduction or even a disappearance of the distance effect. So, in the study by Holloway and Ansari (2009) and that by Kaufmann and Nuerk (2008) the children exhibited normal performance in arithmetic tasks and were confronted with a rather simple single-digit magnitude comparison task they could easily solve. As no backup strategies are necessary in such simple experiments in normally developing children, a larger distance effect was observed to be associated with worse access to the mental number line and – consequently – worse performance in complex arithmetic task. In contrast, consider now the studies by Rousselle and Noël (2007) as well as the present study. As laid out above, Rousselle and Noël (2007) argued that some dyscalculic children may have used back-up strategies even for simple single-digit number comparison. These back-up strategies increased error variance with respect to the distance effect and thus, the distance effect decreased (see Figure 2). Comparably, in the current study the comparison task involved two-digit numbers, of which not all have been taught at school by the time children were assessed in first grade. Nevertheless, most children could do the task. However, for such a complex task even normally developing children may have resorted to back-up strategies similar to that employed by dyscalculic children in more simple tasks. Notably, Nuerk et al. (2004a) report similar findings for children attending grades two to five. In particular, children with a higher error rate (in the number comparison task) were found to exhibit a smaller (decade) distance effect than children with a comparably lower error rate (see Fig. 4). Following the above argument, the children relying on backup-strategies may have boosted error variance and thereby, the distance effect diminished or even disappeared. Put differently, those children able to produce an analogue distance effect in first grade could represent two-digit numbers in an analogue way without major problems (or back-up strategies) and such early magnitude representation of two-digit numbers may corroborate future arithmetic development. To summarize, we suggest that the seemingly contradictory data can be integrated if one postulates a curvilinear relation between distance effect and arithmetic performance which relies on both, explained magnitude variance and disturbing error variance with the latter depending on interactions of task difficulty and individual capability (and especially the need for backup-strategies).

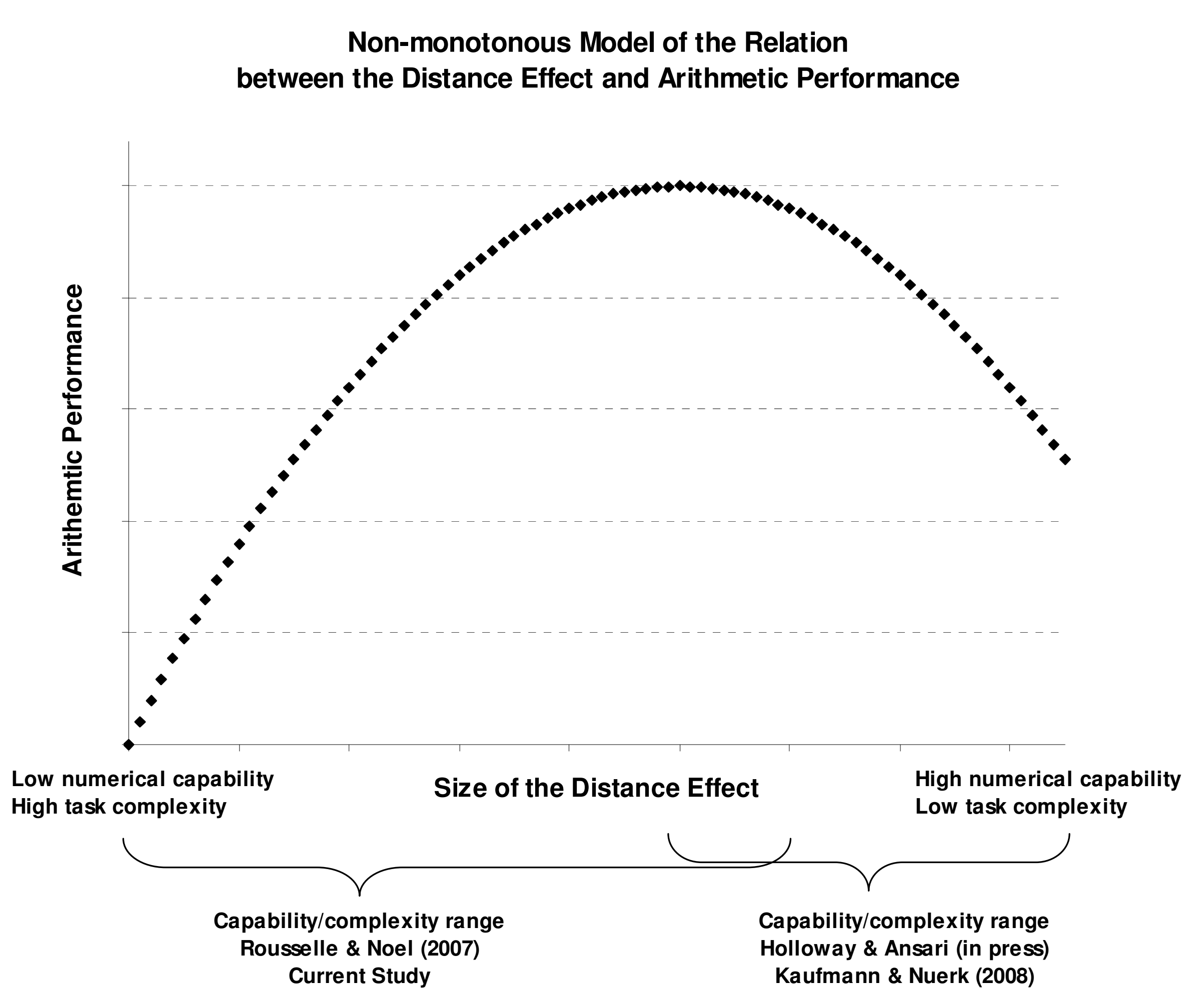


Figure 2: *A hypothetical curvilinear model of a non-monotone relationship between the distance effect and arithmetic performance (see text for details). It is suggested that the size of the distance effect might be determined by two factors: First, increasingly precise and automatic access to the magnitude representation (i.e., higher capability) decreases the explained variance of the distance effect. Second, the distance effect also depends on the error variance which is increased by backup strategies (e.g., counting) necessary for tasks of very high complexity. However, too much of such non-systematic error variance may diminish the distance effect or even make it disappear. The model in Figure 2 accounts for the seemingly antidromic correlations of the distance effect and arithmetic capabilities by assuming that the numerical capability as well as task complexity differ between the cited studies. In Rousselle and Noël's (2007) study, capability is very low (dyscalculic children) while in our study relative task complexity is very high (two-digit number comparison for first graders). In this capability/complexity range, a higher distance effect is particularly associated with less error variance (e.g., less backup strategies) so that a larger distance effect indicates better numerical competencies and better arithmetic performance. Other studies (e.g., Holloway & Ansari, 2009) employed a less complex task (single-digit comparison) and the participants should have had higher individual capabilities (as older children were assessed). For such individually simple tasks, no backup strategies (which boost error variance) are to be expected in normally developing children. Therefore, a smaller distance effect might then indicate more efficient access to the magnitude representation leading to a negative relation between the distance effect and arithmetic performance in this capability/complexity range. Please note that this model remains to be tested empirically, however, at the moment it offers a possible explanation for seemingly contradictory results.*

Working memory and later arithmetic performance

There was one other somewhat unexpected finding in our data. CE and visuo-spatial WM capacity were positively correlated with the carry effect or the error rate for carry problems, respectively. This was surprising and needs further explanation because usually WM capacity in negatively correlated with addition performance in adults (e.g., Imbo et al, 2007c).

In our view there are at least two possible explanations accounting for these unexpected results. (i) Zuber et al. (2009) observed transcoding performance to be reliably moderated by WM capacity in a sample of first graders from which the sample of the present study was recruited. Therefore, possible influences of WM capacity on addition performance in the present study may have been driven by its interrelation with transcoding performance. This interpretation is backed by the current results as a measure of first grade WM span was incorporated in the regression model predicting the carry effect (i.e., CE) and the model accounting for performance for carry addition problems (i.e., visuo-spatial WM span). In this context, it may be important to note that the raw visuo-spatial correlation with later addition performance is not significant and that visuo-spatial WM only became a predictor due to its negative correlation with pure inversion errors in transcoding: Thus, better performance in first grade transcoding (i.e., fewer errors) is associated with better WM (see also Camos, 2008; Zuber et al., 2009). So, the beneficial influences of WM may already be partially considered by the variance explained by the numerical predictors. The remaining influence may be attributed to WM serving as a suppressor variable; by suppressing irrelevant variance of the numerical predictors this resulted in a higher overall R^2. (ii) Another possible reason why we did not obtain reliable influences of WM on addition performance may be that unlike Imbo et al. (2007a) the addition task employed in the current study required choosing the correct result from two presented solution probes rather than actively producing the correct result. Hence, partial results necessary to be kept in mind when trying to compute the result could have been verified by comparing unit and/or decade digit of a preliminary result to the corresponding digits of the solution probes. However, it may have been the case that children with a good WM might not have tried to bypass the addition task by checking the plausibility of the probes unit and/or decade digit but may have tried to calculate the correct result before pressing the corresponding response button. As this reflects the more difficult and thus more error prone strategy these children may have committed more errors which in turn would account for the negative correlation of WM measures with addition performance.

Basic numerical competencies as precursors of future performance

First evidence indicating the relevance of basic numerical competencies as precursors of later arithmetic achievement comes from cross-sectional evaluation of typically developing children (Holloway & Ansari, 2009; see also De Smedt et al., 2009b; Landerl & Kölle, 2009) as well as from dyscalculia intervention (Kaufmann et al., 2003). The current results fit nicely with both lines of evidence identifying basic numerical concepts such as symbolic/non-symbolic magnitude representation or counting principles to serve as fundamentals for later arithmetic achievement. However, as previous studies have only used single digit stimuli in basic numerical tasks, they have not been able to address the issue of place-value integration.

In this way, the current study extends the findings of Holloway and Ansari (2009) in two important aspects: (i) Here, it was observed that when evaluating the predictive power of performance in a magnitude comparison task effective magnitude representation (as reflected by the distance effect) is only part of the story. Rather, it was found that successful identification and comparison of the single digit magnitudes of tens and units complying with the place-value structure of Arabic numbers (as indexed by the unit-decade compatibility effect) is of crucial importance for the further development of arithmetic competencies as well. Please note that in line with this argument Landerl and Kölle (2009) observed the compatibility effect to be reliably more pronounced in children with dyscalculia than in a typically developing sample of control children. This suggests that poor integration of tens and units into the place-value structure of the Arabic number system may be a key symptom of developmental dyscalculia. (ii) The present study was the first to evaluate the developmental trajectories of basic numerical competencies and their influence on arithmetic performance in a longitudinal approach. Thereby, these are the first data that clearly indicate the prominent role of an understanding of basic numerical concepts as a precursor of arithmetic capabilities in typical numerical development.

A second line of evidence corroborating the importance of basic numerical competencies for later arithmetic capabilities comes from recent intervention studies. Dowker (2001) proposed a modular intervention program that incorporated a distinct module covering "understanding the role of place-value in number operations and arithmetic" (Dowker, 2001, p. 7) despite other modules on counting principles, written symbolism for numbers and arithmetical estimation. More recently, Kaufmann and colleagues (2003) picked up on this and aimed to investigate possible differential influences of basic numerical concepts such as counting, transcoding and place-value understanding as well as arithmetic fact knowledge, and procedural knowledge in an intervention programme training dyscalculic children. Post interventional evaluation of children's performance on corresponding numerical tasks revealed that it was the training of basic numerical competencies from which dyscalculic children particularly benefitted. Based on these findings Kaufmann et al. (2003) argue that training of basic numerical knowledge should be a key part of dyscalculia intervention. However, even these intervention results do not automatically qualify the generalisation of basic numerical competencies serving as developmental precursors for arithmetic competencies in typically developing children. In this context, the results of the current study are meaningful. In a longitudinal approach we observed basic numerical competencies and in particular place-value understanding in first grade to be a reliable predictor of arithmetic ability in third grade. Thereby, the present study not only provides first evidence that the interrelation between early basic numerical competencies and later arithmetic capabilities as observed in dyscalculia intervention may also hold for typical numerical development. More particularly, the present results indicate that among basic numerical competencies mastery of the place-value structure of the Arabic number system is of exceptional relevance when it comes to narrowing down precursor competencies of future arithmetic capabilities. However, we wish to make explicit that although the results of the current study point to a crucial role of place-value understanding in the development of numerical capabilities, we would not claim that there are no other basic numerical competencies that serve as a building block of later arithmetic achievement.

Implications for models of numerical development

Interpreting the present results within the framework of current developmental models of numerical cognition remains at a rather hypothetical level as current models of numerical development are mostly theoretical in nature, yet awaiting empirical validation. For instance, the developmental model of numerical cognition proposed by von Aster and Shalev (2007) suggested a four-staged development. In this model the authors attempted to link specific numerical/calculation abilities (arising at consecutive developmental stages) to the relevant brain areas subserving these abilities, on the one hand, and to the cognitive representations underlying these capabilities, on the other hand. Interestingly, while the first two stages (i.e., core magnitude system and verbal number system) are assumed to be acquired by the time of school enrolment acquisition of the third and fourth stage (i.e., Arabic number system and the ordinality of the mental number line) are conceptualized as separate and provisionally later developmental stages to be determined by schooling (see von Aster & Shalev, 2007, Figure 1). Though place-value understanding is not explicitly conceptualized within this model (and to the best of our knowledge in no other neurocognitive model of numerical development), it seems plausible to speculate that place-value understanding emerges at stages three and four (or in later – yet unspecified – stages) of the von Aster and Shalev (2007) model. This assumption is based on the consideration that at least implicit place-value knowledge is a crucial prerequisite for appreciating the numerical value of one-digit and more importantly multi-digit numbers which in turn paves the way for development and the formation of a linear representation of numerical magnitude along a mental number line reflecting also the base-10 property of the Arabic place-value structure (e.g., Moeller et al., 2009a).

With respect to non-numerical domain-general capacities influencing numerical development, the developmental model by von Aster and Shalev (2007) considers – beyond visuo-spatial skills – also WM to be relevant. Indeed, there is accumulating evidence that WM impacts on mathematics performance (e.g., Passolunghi & Mammarella, in press; see Kaufmann & Nuerk, 2005 for an overview) and that WM even seems to be a good predictor of later mathematics achievement (e.g., De Smedt et al., 2009a). Therefore, examining (as well as controlling for) the impact of WM and its components on different aspects of number processing and calculation seems crucial when interested in the developmental trajectories of particular numerical properties such as place-value understanding. Based on such considerations WM measures have been incorporated in the present study (see above).

Although we acknowledge that (and present) empirical evidence supporting theoretical ideas of quite circumscribed areas of numerical development exists (e.g., number magnitude representations: Siegler, 1996; transcoding: Barrouillet et al., 2004; Power & Dal Martello, 1997), we wish to emphasize that there is currently no comprehensive and empirically validated developmental model of numerical development of multi-digit number processing and of the development of place-value integration processes, in particular. Moreover, as already mentioned above, mathematics competency not only depends on well established basic and more complex number processing and calculation skills but rather, flawless mathematics performance requires the interplay of elaborate numerical and non-numerical skills (such as WM for instance). Regarding numerical skills, most research to date was targeted at the representation of number magnitude. Hence, the present study goes one step further by systematically examining developmental trajectories of place-value understanding (being key to complex written arithmetic) in elementary school children.

Therefore, we believe that this study is important in that it adds constraints for further refinement of developmental models of multi-digit (place-value integrated) numerical processing.

CONCLUSIONS

In the current study we investigated the influence of basic numerical competencies, in particular, place-value understanding, on later arithmetic performance in a longitudinal approach. Our results clearly indicate that successful mastery of the place-value concept in first grade is a reliable precursor of arithmetic capabilities in third grade. In this way, the present study extends previous findings on the importance of basic numerical knowledge on the development of arithmetic capabilities in three aspects. (i) This is the first longitudinal observation of an interrelation between basic numerical knowledge and later calculation performance in a sample of typically developing children. (ii) Among basic numerical competencies, we pinpointed early place-value understanding to be of major relevance for later arithmetic abilities in general. (iii) More particularly, we were even able to show that specific numerical effects attributed to a single representation in first grade specifically predict effects related to the same underlying representation in an arithmetic task two years later.

The conclusions for future research are then straightforward. Multi-digit number processing and in particular, the importance of place-value understanding should – in our view – receive greater attention as a unproblematic development of place-value knowledge cannot be taken for granted in all children (e.g., Gervasconi & Sullivan, 2007). Instead, early deficits in place-value understanding still exert their influence on later more complex arithmetic processes. This calls for research not focusing on single-digit numbers processing but which also incorporates representations and concepts necessary to successfully process multi-digit numbers. However, such an approach should not only be restricted to basic research. The present data rather suggest that more specific teaching and training of place-value understanding in education contexts may endorse children's numerical development.

APPENDIX A: Correlation tables of step three analyses evaluating the intercorrelations of specific numerical effects with the carry effect (Panel A), the number of errors for carry problems (Panel B) as well as the number of errors for non-carry problems (Panel C)

A	Task/Effect	1.	2.	3.	4.	5.	6.	7.	8.
Numeric	1. Carry effect (3rd grade)	-	.05	- .01	.14	.09	- .07	.12	.20*
	2. Distance effect (1st grade)		-	- .04	.05	- .19(*)	- .19(*)	.03	- .07
	3. Compatibility effect (1st grade)			-	.13	.04	.04	- .01	- .21*
	4. Inversion (1st grade)				-	.01	- .19(*)	- .16	- .25*
Non-numeric	5. Intelligence (1st grade)					-	.00	- .04	.21*
	6. Verbal WM (1st grade)						-	.27**	.14
	7. Spatial WM (1st grade)							-	.38**
	8. CE (1st grade)								-

B	Task/Effect	1.	2.	3.	4.	5.	6.	7.	8.
Numeric	1. Carry errors (3rd grade)	-	- .13	.23*	.38**	.08	- .02	.14	.02
	2. Distance effect (1st grade)		-	- .04	.05	- .19(*)	- .19(*)	.03	- .07
	3. Compatibility effect (1st grade)			-	.13	.04	.04	- .01	- .21*
	4. Inversion (1st grade)				-	.01	- .19(*)	- .16	- .25*
Non-numeric	5. Intelligence (1st grade)					-	.00	- .04	.21*
	6. Verbal WM (1st grade)						-	.27**	.14
	7. Spatial WM (1st grade)							-	.38**
	8. CE (1st grade)								-

C	Task/Effect	1.	2.	3.	4.	5.	6.	7.	8.	9.
Numeric	1. Non-carry errors (3rd grade)	-	- .17(*)	.23*	.21*	.31**	- .03	.05	.00	- .20(*)
	2. Distance effect (1st grade)		-	- .04	.05	.02	- .19(*)	- .19(*)	.03	- .07
	3. Compatibility effect (1st grade)			-	.13	.39**	.04	.04	- .01	- .21*
	4. Inversion (1st grade)				-	.42**	.01	- .19(*)	- .16	- .25*
	5. Combination errors (1st grade)						- .17	- .03	- .28**	- .37**
Non-numeric	6. Intelligence (1st grade)						-	.00	- .04	.21*
	7. Verbal WM (1st grade)							-	.27**	.14
	8. Spatial WM (1st grade)								-	.38**
	9. CE (1st grade)									-

(*) = $p < .10$; * = $p < .05$; ** = $p < .01$

Section 4

The neurocognitive underpinnings of place-value integration

Study 5

All for one but not one for all: How multiple number representations are recruited in one numerical task

ABSTRACT

Mental arithmetic and calculation rely on a complex network of multiple numerical representations. Usually the components of this network are examined in a between-task approach, which presents the disadvantage of relying upon different instructions, tasks and inhomogeneous stimulus sets across experimental conditions. A within-task approach may avoid these disadvantages and may access the numerical representations more specifically. In the present study we employed a within-task approach to investigate the numerical representations activated in the number bisection task (NBT) in a parametric rapid event-related fMRI study. The only required instruction was to judge whether the middle number of a triplet was also their arithmetic mean (23_26_29) or not (23_25_29). Activity in the left inferior parietal cortex was associated with the deployment of verbal number (fact) representations, while activation of the intraparietal cortex was associated with deeper magnitude processing, instrumental aspects of calculation and activation of the base-10 structure of two-digit numbers. These results replicate strong evidence from the literature. Furthermore, activation in the dorsolateral and ventrolateral prefrontal cortex reveals mechanisms of feature monitoring and inhibition and allocation of cognitive resources in order to solve a specific triplet. We conclude that the network of numerical representations should be studied in a within-task approach and not only with between-task approaches.

INTRODUCTION

Dealing with and the processing of numbers is a highly complex ability relying on multiple cognitive representations such as numerical magnitude (Dehaene, Piazza, Pinel, & Cohen, 2003; Hubbard, Piazza, Pinel, & Dehaene, 2005), arithmetic fact knowledge (Dehaene et al., 2003; Delazer et al., 2003), and procedural knowledge (Butterworth, 2005; Delazer et al., 2004). These different number representations could recently be associated with the activation of specific neuroanatomical structures. Typically, number magnitude processing activated the intraparietal cortex, bilaterally (e.g. Eger, Sterzer, Russ, Girald, & Kleinschmidt, 2003; Dehaene et al., 2003) while arithmetic fact knowledge was associated with activation in the left inferior parietal cortex, in particular the left angular gyrus (Cohen, Dehaene, Chochon, Lehéricy, & Naccache, 2000; Dehaene et al. 2003; Delazer, et al., 2003). However, when calculation problems become more complex, apart from magnitude processing and arithmetic fact knowledge, procedural knowledge coordinating all required processes is necessary (Delazer et al., 2003; Semenza, 2004). It usually subsumes abilities as to select the appropriate strategies and production rules as well as integrating all information needed to solve the task at hand (Delazer et al., 2003). Such procedural knowledge was recently associated with prefrontal and basal ganglia activation (Delazer et al., 2004).

To date, studies investigating the neural correlates of those different number representations and procedural knowledge used a large variety of tasks and materials (Chochon, Cohen, van de Mortele, & Dehaene, 1999; Cohen et al., 2000; Delazer et al., 2004; Fullbright et al., 2000; Goebel, Johansen-Berg, Behrens, & Rushworth, 2004; Menon, Rivera, White, Glover, & Reiss, 2000; Pinel, Dehaene, Riviére, & Le Bihan, 2001; Pinel, Piazza, Le Bihan, & Dehaene, 2004, Simon, Mangin, Cohen, Le Bihan, & Dehaene, 2002; Stanescu-Cosson et al., 2000; Wood, Nuerk, & Willmes, 2006). Due to this diversity of experimental procedures and task requirements, a comparison of brain activation patterns between different studies is very difficult. In many cases, specific numerical processing such as magnitude manipulations cannot be distinguished from more general processes (e.g. response selection, see Goebel et al., 2004). In these cases, differences in brain activation patterns associated with either procedural or fact knowledge, or magnitude representations cannot definitely be attributed to the critical involvement of these constructs, but also to task requirements and stimuli used. Therefore, employing a within-task approach instead of a between-task approach to examine the interactions between procedural and fact knowledge, as well as magnitude representation would be desirable.

Addressing this issue, Nuerk, Geppert, van Herten, and Willmes (2002) presented a modified version of the number bisection task (NBT) as initially used by Dehaene and Cohen (1997) to assess spared numerical abilities in an acalculic patient. In Nuerk et al.'s (2002) verification version of the NBT, healthy adult participants were to determine whether the central number of a triplet also represents the arithmetic mean of the interval (e.g. 23_25_27) or not (e.g. 23_25_29; see also Hoeckner et al., 2008 for results in hemineglect patients). As performance in the NBT is supposed to rely exclusively on magnitude manipulations, the NBT is commonly used to assess number magnitude representation (Cohen et al., 2000; Dehaene and Cohen, 1997). However, Nuerk et al. (2002) showed that different number representations specifically influenced speed and accuracy in the NBT. In their factorial design, Nuerk et al. (2002) identified multiplicativity (i.e. whether or not a triplet was part of a multiplication table) and bisection range (i.e. distance between the outer numbers) to be the most important predictors of performance for correctly bisected triplets. Whereas for triplets not correctly bisected, distance from the central number to the arithmetic mean of the interval and bisection possibility (i.e. whether or not the interval can be bisected by an integer) reliably predicted behavioural performance.

Additionally, Nuerk and colleagues (2002) conducted a regression analysis on item RT to investigate the impact of item properties not manipulated in their stimulus set. Thereby, decade crossing and measures of problem size were identified to generally influence behavioural performance in the NBT: First, classification was systematically faster and less error prone for triplets staying within the same decade (e.g. 23_26_29 vs._25_28_31), and second for triplets with a relatively smaller problem size (e.g. 11_14_17 vs. 81_84_87). In summary, the results of Nuerk and co-workers (2002) suggest that apart from a crucial involvement of magnitude manipulations, also numerical information from other sources (e.g. multiplicativity, procedural rules) is monitored and recruited where beneficial.

For the present study, these six numerical determinants of behavioural performance in the NBT have been selected to investigate their neuronal correlates. Three neural networks incorporating different aspects of these numerical determinants can be distinguished: first, a fronto-parietal network including the intraparietal cortex bilaterally, involved in magnitude processing. Second, the superior frontal and inferior parietal cortex, including the left angular gyrus, which subserves monitoring and applying of procedural rules as well as inhibiting cognitive sets, and finally, the right ventrolateral prefrontal cortex subserving cognitive set changes. Following this differentiation, predictions on the impact of the six determinants introduced above on the activation within each of the three neural networks can be derived:

Range: A large bisection range should lead to a stronger fMRI signal in the intraparietal cortex as more difficult magnitude manipulations are required. Additionally, activation is expected to extend to the posterior superior parietal lobule due to more large-scale navigation on the mental number line (Dehaene et al., 2003). Probably, activation in prefrontal regions as well as in the SMA and pre-SMA regions may be observed, which can be associated with task difficulty (Garavan, Ross, & Stein, 1999). In contrast, when bisection range is small, the intraparietal cortex should be less engaged. Moreover, activation in the inferior parietal lobule, particularly in the left angular gyrus should increase as smaller, less complex problems were to be solved (Qin et al., 2004).

Problem size: Since the frequency of occurrence of a number decreases as its magnitude increases (Dehaene and Mehler, 1992), the representation of relatively smaller numbers is assumed to be more precise and solid (Dehaene, 2001). The angular gyrus is expected to be activated more strongly for triplets with a small problem size. In contrast, when number magnitude increases, deeper magnitude processing is required and thus increased activation in the intraparietal cortex is expected.

Decade crossing: Decade crossing reflects the interaction between magnitude processing and the structure of the symbolic base-10 Arabic system (Wood et al., 2006). When there is no decade crossing, no such symbolic transformations need to be performed, thus activation in the intraparietal cortex should less pronounced. Similar to small bisection ranges an increase in fMRI signal in the left angular gyrus is expected, as familiarity is enhanced when all three numbers of a triplet are from the same decade and, actually, only the unit digits of the triplet have to be bisected.

Distance to the mean: Triplets with a small distance to the correct mean of the interval should lead to increased activation in the intraparietal cortex compared to triplets with a large distance to the mean. This may be attributed to more fine-grained magnitude processing required to differentiate the central number of the interval from its correct mean when these are numerically close to each other.

Multiplicativity: Multiplicatively related triplets should activate the left angular gyrus more strongly than triplets that are not part of a multiplication table indicating retrieval processes for well-learned arithmetic problems (Delazer et al. 2004, Cohen et al., 2000). In contrast, for multiplicatively unrelated triplets, more intense magnitude processing and manipulation is required. Therefore, activation in prefrontal and intraparietal regions should be increased (cf. Delazer et al., 2004).

Bisection possibility: Finally, evaluating bisection possibility might reflect a cognitive set change. When the interval cannot be bisected by an integer, the central (integer) number cannot be the correct response. Therefore, in these triplets all subsequent magnitude manipulations should be discontinued. It is expected that brain activation for triplets impossible to bisect by an integer should specifically activate regions in prefrontal cortex which are involved in cognitive set changes. Whereas for triplets possible to bisect increased activation of the intraparietal cortex would indicate reliance upon magnitude manipulations to derive the correct answer.

METHOD

Participants: Seventeen male right-handed volunteers (mean age=24.2 years, range 21-30 years) took part in the study after having given their written consent in accord with the protocol of the local Ethics Committee of the Medical Faculty.

Experimental task: 360 one- and two-digit number triplets including integer numbers ranging from 1 to 99 were presented in a 2 x 2 within participant design in Arabic notation for both bisected (e.g. 23_26_29) and non-bisected triplets (e.g. 23_25_29). For bisected triplets, multiplicativity (number triplets part/not part of a multiplication table, e.g. 21_24_27 vs. 22_25_28) and bisection range (the distance between the outer numbers; small: 4-8, e.g. 23_26_29 vs. large 12-18, e.g. 22_30_38) were varied. For non-bisected triplets, we manipulated bisection possibility (whether the mean of the interval is an integer, e.g. 21_24_29, mean = 25, or not e.g. 21_24_28, mean = 24.5) and distance of the second number to the mean of the interval (far: 2-8, e.g. 6_7_18 vs. near: 0.5 – 1.5, e.g. 6_11_18).

Problem size, overall magnitude, average parity, parity homogeneity, decade crossing and inclusion of multiples of ten were matched between the respective stimulus groups. In addition to these variables, bisection range and size of the central number relative to the correct mean (i.e. larger/smaller) were matched for non-bisected triplets.

Procedure: Participants were lying in the fMRI scanner while responding to the stimuli. The stimuli were rear-projected on a screen via a mirror attached to the head coil of the scanner at a distance of 12 cm from the participants. Participants were instructed to decide as quickly and accurately as possible whether the interval between the two outer numbers was correctly bisected by the second number (right hand key) or not (left hand key). Each triplet was presented for 3 seconds, followed by an ISI of 1.5 seconds in which only the fixation marks (_ _) were visible. RTs longer than 3 seconds were not recorded. After every eighth trial, there was a longer ISI of 5.6 seconds. All stimuli were presented in Courier New 150. 360 triplets were presented in 5 blocks, each lasting approximately 6 minutes. Between blocks, there was a break of approximately 3 minutes. Trial order was pseudo-randomized so that each stimulus condition appeared about equally often in each block. The experiment was preceded by 30 warm-up trials.

fMRI acquisition: Five functional imaging runs sensitive to blood oxigenation level-dependent (BOLD) contrast were recorded for each participant in a Philips 1.5T Gyroscan MRI system (T2*-weighted echo-planar sequence, TR = 2500 ms; TE = 50 ms; flip angle = 90°; FOV = 220 mm, 64 x 64 matrix; 30 slices, voxel size = 3.4 x 3.4 x 4 mm). Five runs of 146 scans each were acquired. In a rapid event-related design, 360 trials were presented in a pseudo-randomized sequence at a rate of 4.5 seconds in 45 small blocks of 8 trials interleaved with null events of 5.6 seconds. The fMRI time series was corrected for movement artifacts and unwarped using SPM2. Images were resampled every 3 mm to a standard template using sinc interpolation to interpret them in Talairach coordinates (Talairach and Tournoux, 1988), and smoothed with a 6 mm Gaussian kernel.

We convolved brain activation separately for bisected and non-bisected triplets with the canonical hemodynamic response function (HRF) and estimated the effects of parametric regressors. For bisected triplets regressors coding for multiplicativity, decade crossing, bisection range, and problem size were entered as parametric predictors of activation. For non-bisected triplets, regressors coding bisection possibility, decade crossing, bisection range, problem size and distance between the central number and the mean of the interval were entered as parametric predictors of activation. The effects of different regressors on the fMRI signal were estimated using the first-order term of the polynomial expansion of the typical

HRF function (Büchel, Holmes, Rees and Friston, 1998). In order to compare the activation produced by each of these coefficients, the parametric regressors were standardized to a mean of 0 and a standard deviation of 1 in each session. In a random-effects second-level analysis, cortical regions showing modulation of signal specifically due to the parametric regressors were evaluated.

RESULTS

Classification performance

Results concerning RT and errors acquired in the fMRI scanner have already been published (Nuerk et al., 2002) and will only be briefly summarised here (see Table 1). To exclude effects produced by responses to just a subset of items (Nuerk et al., 2001; Nuerk et al., 2002; see also Clark, 1973), we also carried out an analysis of variance (ANOVA) over items (F_2).

ANOVA: Response latencies were shorter for the evaluation of bisected triplets being part of a multiplication table (150 ms; $F_1(1, 16) = 97.52$; $p < .001$; $F_2(1, 176) = 14.59$; $p < .001$) as well as for triplets spanning a relatively smaller range (410 ms; $F_1(1, 16) = 109.36$; $p < .001$; $F_2(1, 176) = 112.46$; $p < .001$). For non-bisected triplets a relatively larger distance between the central number of the interval and its correct mean (256 ms; $F_1(1, 16) = 197.35$; $p < .001$; $F_2(1, 176) = 37.77$; $p < .001$) or the interval being impossible to be bisected by an integer led to faster reaction times (RT) (49 ms; $F_1(1, 16) = 8.41$, $p < .01$; $F_2(1, 176) = 1.81$, $p = .18$). In particular, this main effect of bisection possibility was driven by a significant effect of bisection possibility on RT for isolated non-bisected triplets (57 ms, $t(16) = 3.2$, $p < .01$), it did not reach significance for non-bisectable triplets preceded by another non-bisectable triplet(-17 ms, $t(16) =0.41$, $p = .68$). Comparison of the effect of bisectability in isolated vs. preceded triplets also did reveal a significant difference ($t(16)=2.26$, $p< .05$).

<u>Table 1:</u> Behavioural performance in the NBT as reported in Nuerk et al. (2002). RT in ms and percentage of errors (with SD in parentehesis).

	Bisectable		Non-Bisectable		
<u>RT in ms</u>	Related	Unrelated		Possible	Impossible
Small range	1468 (148)	1594 (140)	Large distance	1629 (116)	1599 (132)
Large range	1853 (140)	2029 (125)	Small distance	1913 (109)	1846 (147)
<u>Errors in %</u>	Related	Unrelated		Possible	Impossible
Small range	3 (3)	4 (4)	Large distance	4 (3)	2 (2)
Large range	10 (7)	20 (10)	Small distance	13 (5)	7 (5)

Regression analysis: RT data were reanalysed in two regression analyses to specifically examine the influence of the above described determinants. Note that the differences between results of the new and the original analysis (see Nuerk et al., 2002, p. 703) are due to the predictors entered in the respective analyses. In the current analysis, range, decade crossing, sum and multiplicativity were included as predictors in the regression model for bisected triplets. For non-bisected triplets, range, decade crossing, sum, bisectability and distance to the mean were included as predictors. Regression models accounted for a considerable proportion of variance in bisectable (R^2=0.52) and non-bisectable RT (R^2=0.51; see Table 2). Range, decade crossing, sum and multiplicativity were significant predictors of RT in bisected triplets. For non-bisected triplets, range, decade crossing, sum, and distance to the mean were significant predictors of RT. Although bisectability was not a significant predictor of RT, we have included it in the general linear model examining fMRI signal in non-bisected triplets, since the main effect of bisectability was significant in the ANOVA.

Table 2: Regresion analysis for bisectable and non-bisectable triplets.

Bisectable triplets (adjusted R^2 = .52)				
Predictor	Standardized beta	t(175)	Raw correlations	Partial correlations
Range	.519	8.89**	.650	.457
Decade crossing	.255	4.37**	.516	.224
Problem size	.132	2.56*	.178	.132
Multiplicativity	-.206	-4.01**	-.219	-.206
Non-Bisectable trials (adjusted R^2 = .51)				
Predictor	Standardized beta	t(174)	Raw correlations	Partial correlations
Range	.351	5.30*	.425	.276
Decade crossing	.441	7.22*	.556	.376
Problem size	.111	2.03*	.111	.106
Bisection possibility	.055	1.04	.086	.054
Distance to mean	-.446	-7.94**	-.258	-.414

* p<.05; ** p<.005, two-sided.

fMRI

Bisection Range, decade crossing and problem size (operationalized as the sum of the three numbers of a triplet) were common predictors in both bisected and non-bisected triplets. The pattern of activation elicited by these predictors was identical for both bisected and non-bisected triplets and will be described jointly. Multiplicativity was a parametric predictor of bisected triplets only, while bisection possibility and distance to the mean were parametric predictors of non-bisected triplets.

Large range > small range: When bisection range increased, fMRI signal increased bilaterally in large portions of the left and right intraparietal cortex, with projections to the extrastriate and striate cortex. Additionally, cortex areas at the left temporo-occipital junction were activated. Furthermore, activation was observed in the left and the right supplementary motor area as well as the premotor cortex, bilaterally. Finally, different clusters of activation were found bilaterally in the dorsolateral and ventrolateral prefrontal cortex as well as in the homologue of Broca`s area (Figure 1A, Table 3).

Small range > large range: When bisection range decreased, increased fMRI signal was found in the anterior cingulate, the right dorsolateral prefrontal cortex, and the retrosplenial cortex (Figure 1A, Table 3).

Decade crossing > non-crossing: Activation in a large network of cortical regions was observed for triplets crossing a decade boundary (Figure 1B, Table 4). The intraparietal cortex as well as areas in the temporo-occipital junction was activated bilaterally as was the extrastriate cortex. Further clusters of activation were found bilaterally in the dorsolateral prefrontal cortex, in the SMA and the premotor cortex. Finally, the ventrolateral prefrontal cortex was activated bilaterally.

No-decade-crossing > decade crossing: Stronger fMRI signal was registered in the angular gyrus as well as in the supramarginal gyrus bilaterally. Further activation clusters were observed in the dorsolateral prefrontal cortex, anterior cingulate and retrosplenial cortex bilaterally (Figure 1B, Table 4).

Large problem size > small problem size: When problem size increased, activation was increased in the striate and extrastriate cortex bilaterally, in the left temporo-occipital junction and medial portions of the cerebellum (Figure 1C, Table 5).

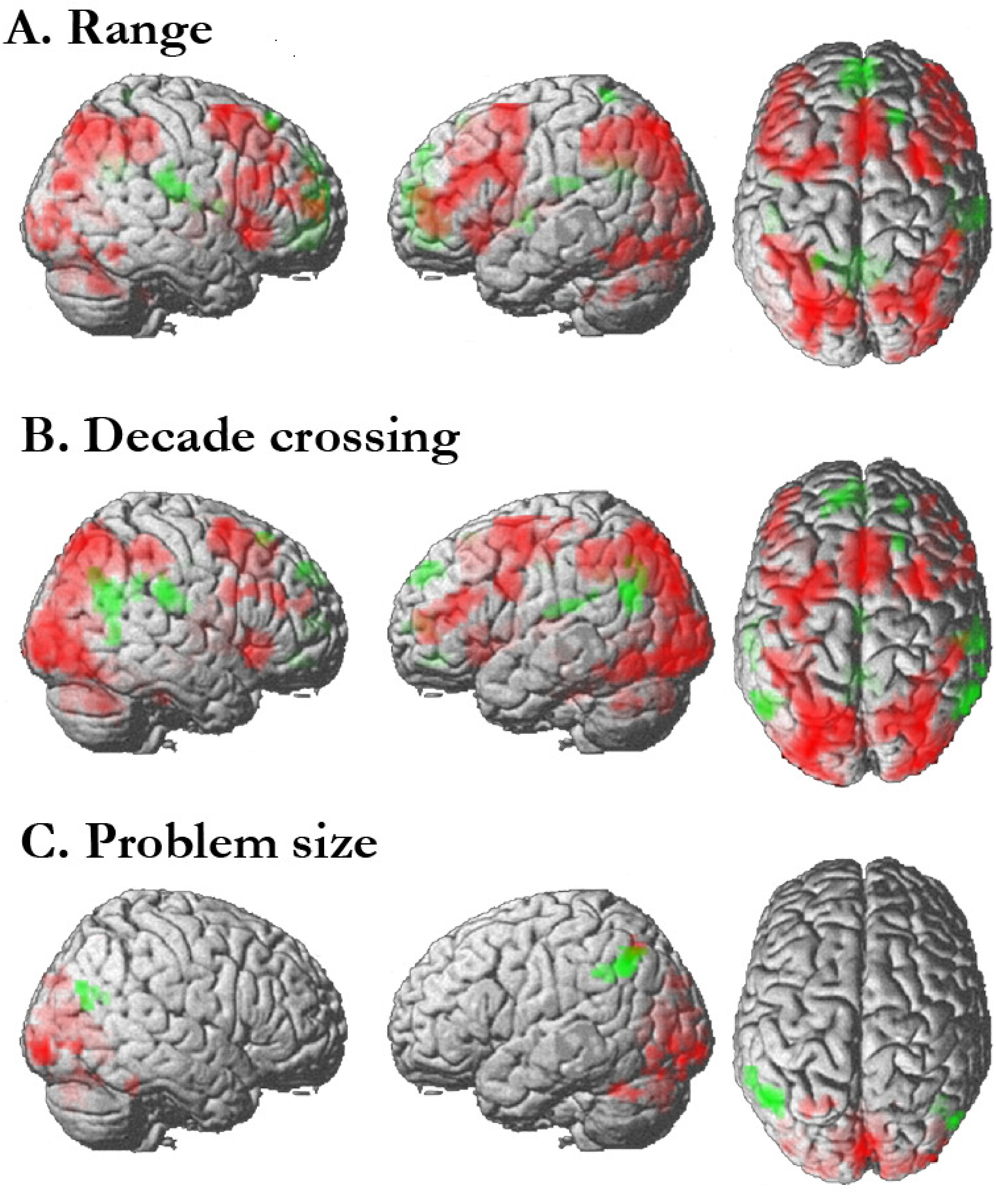


Figure 1: *fMRI signal for parametric predictors common to bisected and non-bisected triplets. All constrasts were thresholded at p = .001, voxelwise uncorrected, and minimal cluster size set to k = 10. 1A: Bisection range. Regions coloured red depict voxels showing stronger signal for large bisection range. Regions coloured green represent voxels exhibiting stronger signal for small bisection range. 1B: Decade crossing. Regions coloured red represent voxels showing stronger signal for triplets with at least one decade crossing, than for triplets without decade crossing. Regions coloured green depict voxels exhibiting stronger signal for triplets without decade crossing than for triplets with at least one decade crossing. 1C: Problem size. Regions coloured red represent voxels showing stronger signal for large problem size than for small. Regions coloured green indicate voxels showing stronger signal for small problem size than for large.*

Table 3: Brain areas activated by bisection range (all p < 0.001, voxelwise uncorrected)

Region	Talairach coordinates x, y, z	t-value df=16	BA	Cluster size k
Large range > small range				
Right intraparietal cortex	50, -33, 49	8.26	40	263
Right intraparietal cortex	35, -53, 52	7.79	7	-
Left posterior intraparietal cortex	-9, -73, 48	8.56	7	276
Left anterior intraparietal cortex	-50, -36, 46	4.85	40	10
Right extrastriate cortex	30, -71, 39	9.26	19	29
Right striate cortex	39, -90, 2	4.75	18	26
Left extrastriate cortex	-27, -71, 39	7.27	19	76
Left temporo-occipital juction	-39, -68, -12	6.31	19/37	50
Left striate cortex	-36, -82, -11	5.26	18	16
Right supplementary motor area	6, 20, 43	10.97	8	44
Left supplementary motor area	-3, 17, 49	10.33	8	29
Right premotor cortex	30, 14, 44	8.44	6	52
Left premotor cortex	-24, 12, 50	7.55	6	104
Right dorsolateral prefrontal cortex	50, 31, 34	5.62	9/46	23
Left dorsolateral prefrontal cortex	-45, 4, 27	8.81	9	110
Right homologue of Broca`s area	42, 20, 2	4.74	45	18
Right ventrolateral prefrontal cortex	39, 17, -6	7.22	47	27
Left ventrolateral prefrontal cortex	-30, 20, -4	5.14	47	19
Small range > large range				
Anterior cingulate	-3, 46, -10	8.06	32	363
Right dorsolateral prefrontal cortex	18, 32, 51	7.52	9	13
Retrosplenial cortex	-3, -45, 33	8.53	7/31	234

Small Problem size > large problem size: The left inferior parietal cortex and to a smaller extent the right inferior parietal cortex exhibited increased fMRI signal as a result of small problem size. (Figure 1C, Table 5). In the left hemisphere, a small cluster of activation was present projecting from the inferior to the superior parietal cortex.

Small distance to the mean > large distance to the mean: When distance to the mean decreased, stronger fMRI signal was observed in the following regions: Left anterior intraparietal cortex, posterior intraparietal and neighbouring extrastriate cortex bilaterally, Broca's area, SMA/pre-SMA, frontal eye-fields and in the frontal operculum bilaterally (Figure 2A).

<u>Table 4:</u> Brain areas activated by decade crossing (all p < 0.001, voxelwise uncorrected)

Region	Talairach coordinates x, y, z	t-value df=16	BA	Cluster size k
Decade crossing > no decade crossing				
Right intraparietal cortex	33, -53, 50	8.52	7/40	250
Right intraparietal cortex	33, -59, 44	7.55	7	-
Left intraparietal cortex	-39, -33, 40	9.49	5/7	327
Left intraparietal cortex	-30, -65, 50	8.00	7	-
Right extrastriate cortex	33, -74, 40	6.87	19	77
Left extrastriate cortex	-27, -74, 37	10.02	19	117
Right temporo-occipital juction	42, -68, -12	6.62	19/37	14
Left temporo-occipital juction	-45, -62, -12	7.66	19/37	52
Right dorsolateral prefrontal cortex	59, 22, 27	5.81	9/46	19
Left dorsolateral prefrontal cortex	-45, 4, 30	7.46	9	42
Left dorsolateral prefrontal cortex	-45, 30, 21	5.83	9/46	38
Right supplementary motor area	6, 17, 43	9.32	8	41
Left supplementary motor area	-3, 17, 49	8.90	8	28
Right premotor cortex	30, 11, 44	6.48	6	56
Left premotor cortex	-35, -9, 47	8.34	6	216
Right ventrolateral prefrontal cortex	39, 17, -6	5.09	47	22
Left ventrolateral prefrontal cortex	-30, 20, -9	4.99	47	13
No decade crossing > decade crossing				
Right angular gyrus	56, -63, 31	5.56	39	102
Left angular gyrus	-53, -60, 28	6.85	39	88
Right supramarginal gyrus	59, -25, 26	7.95	40	78
Left supramarginal gyrus	-62, -25, 18	5.38	40	28
Right dorsolateral prefrontal cortex	21, 48, 34	5.92	9	11
Left dorsolateral prefrontal cortex	-12, 51, 31	5.72	9	49
Right retrosplenial cortex	9, -48, 27	5.87	31	64
Anterior cingulate gyrus	0, 43, -12	6.70	32	35

Large distance to the mean > small distance to the mean: When distance to the mean increased, stronger fMRI signal was registered in the inferior parietal cortex and the anterior portion of the superior frontal gyrus, both bilaterally. A further activation cluster was observed in the right medial temporal gyrus (Figure 2A, Table 6).

Possible > impossible: No region exhibited an increase in activation for this contrast at the voxelwise uncorrected threshold of $p = .01$.

Impossible > possible: Increased activation in the right ventro-lateral prefrontal cortex was observed for triplets impossible to bisect by an integer (BA47, Talairach coordinates: 40, 20, -4, $k = 10$, $t(16) = 5.99$, $p < .001$, uncorrected, see Figure 2B and Table 6).

Multiplicative > non-multiplicative: When the triplet was part of a multiplication table, the left angular gyrus and a small portion of prefrontal cortex was specifically activated (Figure 3, Table 7).

<u>Table 5:</u> *Brain areas activated by problem size (all p < 0.001, voxelwise uncorrected)*

Region	Talairach coordinates x, y, z	t-value df=16	BA	Cluster size k
Large sum > small sum				
Right extrastriate cortex	33, -90, -3	6.28	19	602
Left strate cortex	-9, -81, 6	5.94	17/18	-
Left extrastriate cortex	-15, -93, 0	5.03	19	-
Right strate cortex	3, -84, 15	5.07	17/18	-
Left temporo-occipital junction	-42, -53, -20	5.82	19/37	28
cerebellum	-6, -74, -19	7.49	-	85
Small sum > large sum				
Left inferior parietal cortex	-48, -62, 47	5.50	39	66
Right inferior parietal cortex	53, -69, 23	5.33	39	26

Non-multiplicative > multiplicative: For triplets not part of a multiplication table, increased fMRI signal was observed bilaterally in the intraparietal cortex, in posterior portions of the middle frontal gyrus, SMA/pre-SMA, and in the left temporo-occipital junction (Figure 3, Table 7).

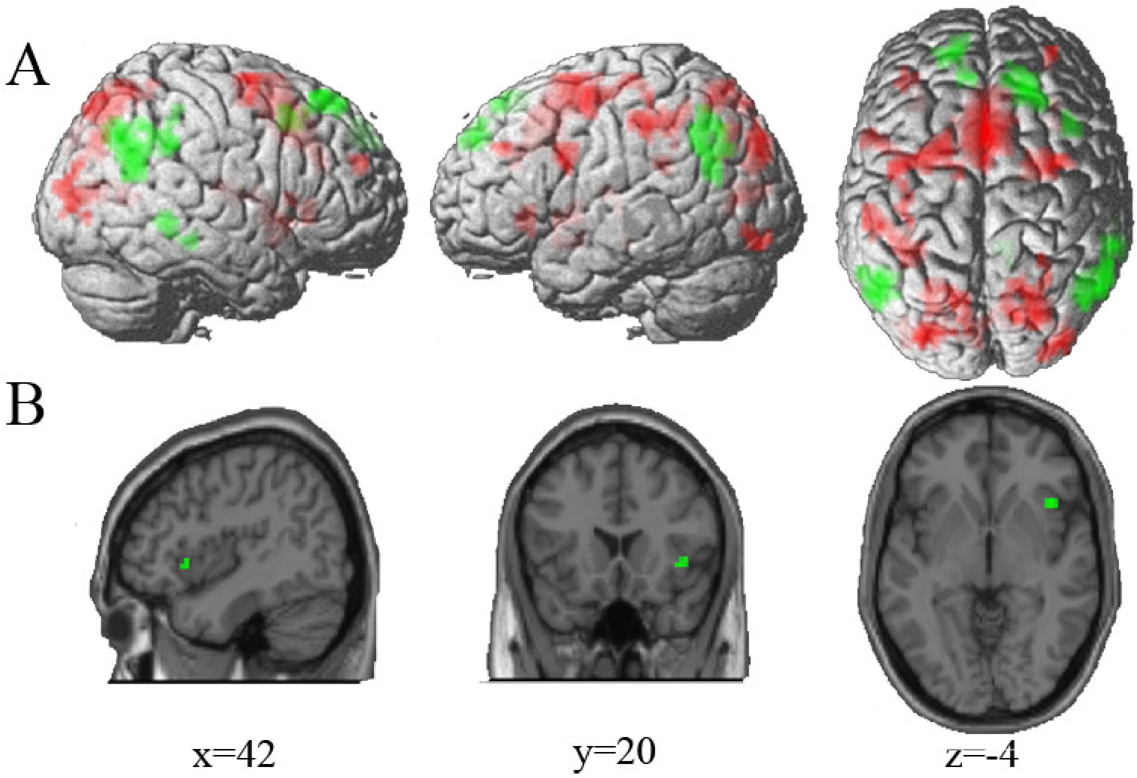
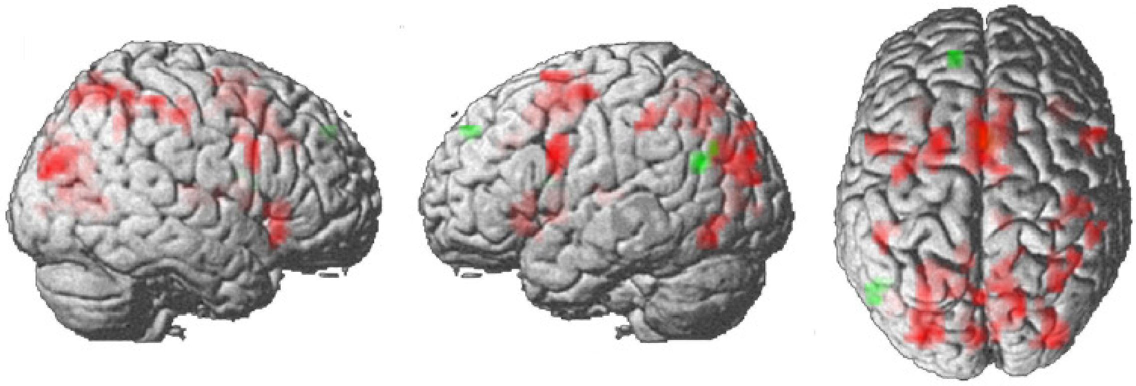


Figure 2: fMRI signal for parametric predictors only included for non-bisected triplets. All contrasts were thresholded at p = .001, voxelwise uncorrected, and minimal cluster size set to k = 10. 2A: Distance to the correct mean. Red coloured regions represent voxels exhibiting stronger signal for larger distances to the mean than for smaller ones. Regions coloured green represent voxels showing stronger signal for smaller distances to the mean than for larger ones. 2B: Bisection possibility. Regions coloured green depict voxels showing stronger signal when the interval was impossible to bisect by an integer.



<u>Figure 3:</u> fMRI signal for parametric predictors only included in bisected triplets. All contrasts were thresholded at p = .001, voxelwise uncorrected, and minimal cluster size set to k = 10. Red coloured regions indicate voxels showing stronger signal for triplets not part of a multiplication table. Green coloured regions represent voxels showing stronger signal for triplets part of a multiplication table.

<u>Table 6:</u> Brain areas activated by distance to correct mean (all p < 0.001, voxelwise uncorrected)

Region	Talairach coordinates x, y, z	t-value df=16	BA	Cluster size k
Small distance to correct response > large distance to correct response				
Left anterior intraparietal cortex	-39, -42, 44	5.73	40	52
Left posterior intraparietal/extrastrate cortex	-21, -56, 50	5.61	7	50
Right posterior intraparietal/extrastrate cortex	15, -55, 58	7.91	7	165
Left dorsolateral prefrontal cortex	-50, 10, 27	5.82	46	50
SMA/pre-SMA	0, 11, 49	11.44	8	467
Right frontal eye field	30, 0, 61	6.44	8	59
Left frontal eye field	-39, -1, 47	5.51	8	169
Left frontal operculum	-30, 27, -3	6.29	45	25
Right frontal operculum	33, 23, 2	5.54	45	18
Left occipital cortex	-39, -75, -1	5.56	19/37	23
Large distance to correct response > small distance to correct response				
Left inferior parietal cortex	-45, -54, 33	7.09	39/40	123
Right inferior parietal cortex	45, -59, 36	6.15	39/40	192
Left dorsolateral prefrontal cortex	-15, 51, 31	6.15	9	40
Right dorsolateral prefrontal cortex	18, 35, 51	8.87	9	69
Right medial temporal gyrus	65, -41, 0	7.03	21	18

<u>Table 7:</u> Brain areas activated by multiplicativity (all p < 0.001, voxelwise uncorrected)

Multiplicative > non-multiplicative				
Region	Talairach coordinates x, y, z	t-value df=16	BA	Cluster size k
Left angular gyrus	-50, -60, 25	5.24	39	17
Left dorsolateral prefrontal cortex	-15, 48, 36	5.31	9	12
Non-multiplicative > multiplicative				
Left anterior intraparietal cortex	-48, -36, 43	5.23	40	21
Right anterior intraparietal cortex	42, -44, 52	5.96	40	24
Left posterior intraparietal/extrastrate cortex	-33, -75, 23	8.12	7/19	201
Right posterior intraparietal/extrastrate cortex	18, -67, 56	8.14	7/19	67
Left dorsolateral prefrontal cortex	-45, 7, 25	5.62	46	74
Right dorsolateral prefrontal cortex	56, 13, 24	4.57	46	22
SMA/pre-SMA	-3, 20, 43	6.50	8	231
Left frontal operculum	-27, 21, 4	4.71	45	29
Right frontal operculum	39, 23, -11	7.09	45	47
Left temporo-occipital junction	-45, -64, -7	5.63	19/37	28

DISCUSSION

As the behavioural data were extensively discussed in Nuerk et al. (2002) the following will concentrate on discussing fMRI activation patterns. Results show that specific numerical representations influence the fMRI signal. Cortical areas activated in the present study are associated with three groups of cognitive processes: the first one includes a bilateral fronto-parietal network subserving magnitude processing, symbolic base-10 processing, and other instrumental aspects of number processing. The second one includes regions in the dorsolateral prefrontal- and inferior parietal cortex, in particular the left angular gyrus, which are associated with monitoring of different sources of information and arithmetic fact knowledge, respectively. The third one involves specific regions in the ventrolateral prefrontal cortex responsible for cognitive set changing. The determinants of the neural activation observed in each one of these neural networks will be discussed in turn.

The fronto-parietal network and the intraparietal cortex

Two components of the activation of the network involving prefrontal and parietal regions can be distinguished: First, the recruitment of general cognitive processes such as working memory, visual attention and response selection were recurrently associated with fronto-parietal activation. Second, specific numerical processes such as magnitude processing (Hubbard et al., 2005) and symbolic base-10 processing (Wood et al., 2006; Hoeckner et al., 2008; Knops, Nuerk, Sparing, Foltys, & Willmes, 2006) were found to activate a fronto-parietal network.

On the one hand, more complex mental arithmetic activates the fronto-parietal network including the intraparietal cortex bilaterally (Kong et al., 2005). Therefore, activation within these neural structures may reflect the deployment of instrumental resources required for more demanding tasks (Gruber, Indefrey, Steinmetz, & Kleinschmidt, 2001). On the other hand, recent research showed that number processing might selectively activate the intraparietal cortex even when controlling for task complexity and response selection demands (Ansari, Dhital, & Siong, 2006a; Ansari, Fugelsang, Dhital, & Venktraman, 2006b; Piazz, Izard, Pinel, Le Bihan, & Dehaene, 2004). Finally, activation in the posterior superior parietal lobule was associated with navigating upon the mental number line (Dehaene et al. 2003) and with symbolic magnitude processing (Wood et al., 2006). Since the activation patterns found in the present study are similar to those found in the studies mentioned above, we suggest that both, general and specific numerical, processes can be associated with activation in the fronto-parietal network. The specific contribution of each of these processes to the overall activation pattern will be discussed.

Magnitude processing

Stronger activation of number magnitude was observed in three contrasts: (i) when triplets spanned a larger number range, (ii) when the distance to the mean was small (applicable only for non-bisected triplets), and (iii) when triplets were not part of the multiplication table (applicable only for bisected triplets).

Increased activation in the bilateral intraparietal cortex was found for triplets spanning a large range. This reflects more demanding magnitude processing in these triplets as magnitude representation becomes less accurate in larger numerical intervals (i.e. according to the Weber-Fechner law, see Dehaene et al., 2003). Activation in the bilateral intraparietal cortex also increased when distance between the middle number and the correct mean was small. So, difficulty in rejecting the middle number was inversely proportional to the distance between the central number and the mean of the two outer numbers. This result can be interpreted as a distance effect for the magnitude comparison between a standard (i.e. the correct mean of the interval) and a probe (i.e. the actual central number of the triplet; Goebel et al., 2004; Pinel et al., 2001, Wood et al., 2006).

Moreover, when triplets were not part of a multiplication table stronger activation of the fronto-parietal network was observed. Again, this increase in activation indicates deeper magnitude processing in these triplets. A trivial item difficulty explanation cannot account for stronger frontoparietal activation in non-multiplicative triplets, since overall magnitude, range and problem size were matched between multiplicative and non-multiplicative triplets. Therefore, when evaluating non-multiplicative triplets, participants rely more on the magnitude representation than in multiplicative triplets. This result is in line with an interpretation for the changes in neuroanatomical regions activated before and after training arithmetical problems (Delazer et al., 2003; 2005; Ischebeck et al., 2006; Ischebeck, Zamariaan, Egger, Scjocke, & Delazer, 2007). In their training studies, participants practiced arithmetical problems for some days. Without practice, brain activation was stronger in the intraparietal cortex. Whereas, after practice the main locus of activation for trained items was in the left angular gyrus. The authors accounted for this change in activation by the formation of lexical representations for items repeatedly presented during training.

Finally, increased activation in the posterior superior parietal lobule was also observed in the contrasts discussed above. Activation in this area is commonly attributed to the activation of visuo-spatial representations, which may reflect mental navigation on a spatially oriented number line (Dehaene et al., 2003).

Symbolic base-10 processing

Decade crossing also activated the fronto-parietal network, particularly the intraparietal cortex bilaterally. When a decade boundary had to be crossed, determining the correct mean of the interval not only seemed to activate the holistic magnitude representation of the involved numbers, but also the decomposed magnitudes of tens and units. As range was matched between triplets with and without decade crossing this effect cannot be attributed to the activation of overall number magnitude, but corroborates the interpretation of decomposed magnitude representations for tens and units (Wood et al., 2006). Recent fMRI studies have shown that the magnitudes of tens and units of a two-digit number are represented separately within the intraparietal cortex (Wood et al. 2006; Knops et al. 2006). The decomposed magnitude representations of tens and units seem to rely heavily upon visuo-spatial representations. For hemineglect patients Hoeckner et al. (2008) observed a larger impairment in triplets with a decade crossing between the first and the central number than in triplets with a decade crossing between the central and the third number in the NBT. This is equivalent to a specific deficit in accessing the magnitude of relatively smaller decade digits represented further to the left of the mental number line (Zorzi, Priftis, & Umiltà, 2002). Therefore, the effect of decade crossing on fMRI signal may indicate a close interaction between the intraparietal sulcus and the posterior superior parietal lobule, which was proposed to be responsible for precise navigation on the mental number line (Dehaene et al. 2003).

Other areas associated with the fronto-parietal network

In general, activation in the intraparietal cortex extended to the adjacent extrastriate cortex BA 19, bilaterally. Activation in the extrastriate cortex seems to be driven by top-down processes originating from the intraparietal cortex. In particular, Weiss, Marshall, Zilles, and Fink (2003) showed that parietal regions exerted top-down regulation over the activation in the extrastriate cortex when visuo-spatial processing became more demanding. Such activation of large portions of bilateral visual cortex accompanying the activation of the intraparietal cortex in tasks involving number magnitude processing has been reported in previous studies (Pinel et al., 2001; Wood et al., 2006). Another area activated in parallel with the intraparietal cortex is the temporo-occipital area, also known as the visual number form area (Cohen et al., 2002; Cohen et al., 2004). Activation in this region, mainly in the left hemisphere, can be explained by more intensive processing of the visual identity of Arabic digits in more complex triplets. Deeper visual processing was again required by the bisection of large numerical ranges, by the computation of small distances to the mean, by non-multiplicative triplets, and by decade crossing.

Finally, the prefrontal area activated in contrasts demanding deeper magnitude processing in the present study overlapped with those regions that Doricchi et al. (2005) associated with the deficit in numerical bisection performance of hemineglect patients. The patients examined by Doricchi, Guariglia, Gasparini, and Tomaiuolo (2005) had more pronounced lesions in the prefrontal cortex than other hemineglect patients. Additionally they were also more impaired in a task of spatial working memory. Thereby, the activation in the dorsolateral prefrontal cortex observed in the NBT may be attributed to spatial working memory processes necessary for bisecting numerical intervals.

In sum, items that are more complex led to increased magnitude processing in the intraparietal cortex. Moreover, via top-down regulation, this activation also spread into the posterior superior parietal lobule as well as other areas subserving visuo-spatial processes.

Inferior parietal cortex, the angular gyrus and fact knowledge

The fMRI signal in the angular gyrus was stronger (i) when numerical range was small, (ii) when problem size was small, (iii) when numbers were part of a multiplication table, (iv) when no decade crossing occurred, and (v) when the distance to the mean increased. Two main reasons for the activation of the angular gyrus shall be discussed: First, the unspecific familiarity with relatively smaller numbers, and second, the capacity to monitor procedural and rule-based information from different sources.

Familiarity with arithmetic problems

Small and more frequently encountered arithmetic problems establish stronger representations. Therefore, they are solved more efficiently than larger and less frequent problems (Domahs, Delazer and Nuerk, 2006). In the present study, stronger fMRI signal was found in the anterior cingulate gyrus and retrosplenial cortex when range was small (Figure 1A). Additionally, when problem size was small (i.e. a triplet consisting of relatively small numbers, e.g. 3_5_8), increased activation was observed in the left angular gyrus and in the right angular gyrus to a smaller extent (Figure 1C). Generally, increased activation in the anterior cingulate gyrus and the retrosplenial cortex is associated with processing more familiar information (Sugiura, Shah, Zilles, & Fink, 2005; Shah et al., 2001). For the case of the NBT this means that activation in the retrosplenial cortex and angular gyrus is associated with the processing of smaller and/or easier triplets. Specifically, these are triplets with a small problem size and spanning a small range. Thus, the higher familiarity with the numbers constituting such triplets can be interpreted as a corollary of the higher exposure to them.
Additionally, activation in the angular gyrus was often accompanied by a small portion of the superior prefrontal cortex being activated (e.g. Figure 3, multiplicative>non multiplicative triplets contrast). Activation in this region may reflect the involvement of executive mechanisms responsible for monitoring different cognitive features of the triplet (Brass and

von Cramon, 2004) and inhibition of a cognitive set (Konishi, Jimura, Asari, & Miyashita, 2003). Such mechanism may monitor the occurrence of information dispensable or redundant for solving the NBT, but which can be useful to solve it more efficiently. We will elaborate more on this point in the next section and relate it to the processing of multiplicativity and distance to the correct mean.

Monitoring numerical information from different sources and inhibition

Activation of the superior frontal gyrus and the angular gyrus have been associated with the monitoring of information from different sources (Brass and von Cramon, 2004), inhibition of a cognitive set (Konishi et al., 2003) and verification of rules (Skosnik et al., 2002), respectively. In the current study, the angular gyrus was activated when numerical distance to the mean was large, when no decade crossing occurred, and when triplets were part of a multiplication table (Figures 1B, 2A and 3, respectively). The superior frontal gyrus bilaterally and the posterior cingulate gyrus, were also activated in these contrasts as well as the supramarginal cortex for the contrasts small distance to the mean > large distance to the mean and no decade crossing > decade crossing. Despite a great disparity between the numerical properties represented in these contrasts, there is a crucial similarity common to all of them. They all convey a procedural rule, which, when correctly applied, were useful for solving a triplet. When the middle number was numerically close to one of the outer numbers – being consequently far from the correct mean- (e.g. 32_33_41), computations might be interrupted, as it is very unlikely that the central number is the correct mean of the interval. Similarly, when no decade crossing occurred, the decade digits might be ignored, since they were all in the same decade (e.g. 92_95_98). Furthermore, when a triplet was part of a multiplication table, the central number is per definition the mean of the interval. Always when these rules can be applied, further computations involving magnitude comparison were unnecessary. However, information about decade crossing and multiplicativity is not primarily necessary for solving the NBT. Therefore, participants need to verify the applicability of these rules for each trial. For that, it is necessary to monitor in parallel the different numerical properties of the triplet as well as to inhibit the magnitude processing. In the present study, participants were able to monitor for multiplicativity, presence of decade-crossing distance from the central number to the mean and number magnitude. Furthermore, part of the activation observed in the superior frontal gyrus may be due to inhibition processes (Konishi et al., 2003) suppressing magnitude manipulations as these are not required in triplets for which the correct response can be inferred by using a procedural rule. For instance, in triplets without decade crossing participants may have inhibited the magnitude of tens because the comparison of unit distances is sufficient for responding correctly. Furthermore, in multiplicative triplets the central number is per definition the correct mean of the interval. Therefore, no magnitude processing would be needed at all. Additionally, in triplets with a large distance to the mean, the central number is numerically too close to one of the outer numbers to be the correct mean of the interval. Again, magnitude manipulations may be discontinued in these cases[1].

In the particular case of the stronger activation in the angular gyrus found for multiplicatively related triplets, a second account for the current data is possible. Apart from the above-mentioned monitoring and inhibitory mechanisms, activation of the left angular gyrus has also been associated with arithmetic fact knowledge (Delazer et al. 2003; 2005). Such an interpretation is also compatible with the data pattern observed for multiplicativity. Therefore, the increased activation in the (left) angular gyrus may also indicate table-related multiplication facts are accessed while performing the NBT (Cohen et al., 2000).

In summary, when numbers keep well-learned relations with each other and these relations are more frequently activated and manipulated, activation in the left inferior parietal cortex, especially in the angular gyrus increased. Moreover, activation of the superior frontal gyrus increased when a triplet could be solved by applying procedural rules. When none of the rules applied, participants needed to determine the mean of the interval and compare it with the magnitude of the central number[1]. However, in the case of non-bisectable triplets the number that correctly bisects the interval is not an integer. Therefore, per definition the central number of the triplet cannot be the correct mean of the interval. In this particular case, there is no need to compare the magnitudes of the central number and the correct mean of the interval. In other words, the procedure of comparing the magnitude of the central number with the correct mean of the interval becomes meaningless in the case of non-bisectable triplets. Changes in the relevance of mental rules are associated with the activation of a specific region in the ventrolateral prefrontal cortex.

Ventrolateral prefrontal cortex

Imaging and patient studies show that the ventrolateral prefrontal cortex is associated with controlled retrieval of mental rule meanings and generation of alternative solutions for a problem (Donohue,Wendelken, Crone, & Bunge, 2005, Miller & Tippet, 1996). Miller and

---

[1] Note that the NBT can be solving using different strategies: participants may for instance subtract the first number from the central, add the result to the central number again and compare it with the third number. Alternatively, participants may calculate the mean of the two outer numbers and compare the result with the central one. In both cases, triplets with small bisection ranges may be solved by recruiting fact retrieval while trials large bisection ranges and/or decade crossing may be more demanding and may result in stronger activation of the fronto-parietal network. However, there are important similarities between the two strategies: in both cases, a distance is computed and compared with a standard. In the first case, the distance to the third number, in the other case the distance to the middle is calculated. Independently of the strategy used, however, in both cases an effect of decade crossing, problem size, range, bisection possibility and distance to mean are expected. Finally, it is possible that the underscore character used to separate the numbers in the NBT induced the use of subtraction strategies in the present study. Since subtraction strategies are associated with magnitude manipulation, the use of procedural rules may be even stronger when the underscore character is not utilized. In follow-up studies, this aspect should be investigated further.

Tippet (1996) have shown that patients with brain lesions extending to the right ventral prefrontal cortex were selectively impaired in tasks involving the generation of alternative solutions for a problem. Furthermore, Goel and Vartarian (2005) reported increased activation in this region when participants successfully generated alternative solutions for a task.

In the present study, the right ventrolateral prefrontal cortex was the only region (see Figure 2B) showing more activation for non-bisectable triplets. These results suggest that the right ventrolateral prefrontal cortex is relevant for checking, whether the comparison between the mean of the interval and the central number should be completed or not. When the mean of the two outer numbers is not an integer number (e.g. for 21_25_30 the mean of 21 and 30 is 25.5), number magnitude comparison should be discontinued, because the central number cannot be the correct response of the bisection. Behavioural data are in line with neurofunctional evidence showing that activation in the ventrolateral prefrontal cortex is specific for cognitive set changes (Goel & Vartarian, 2005). While responses to non-bisectable triplets presented in isolation were significantly faster than to bisectable triplets, response latencies for non-bisectable triplets following another non-bisectable triplet did not differ from that for bisectable triplets. This increase in reaction time may reflect costs imposed by again activating the original cognitive set which specified a comparison between the mean of the interval and the central number.

Since right ventrolateral prefrontal cortex is engaged in the rational use of cognitive resources and cognitive set changes, we suggest that this area may be considered part of the neural system subserving procedural and conceptual knowledge about numbers and their manipulations.

CONCLUSIONS

The results of the present study strongly suggest that different neurocognitive systems dedicated to (i) magnitude and symbolic base-10 processing, (ii) familiarity, monitoring and inhibiting irrelevant information and fact knowledge, and (iii) cognitive set changes, are selectively activated in the number bisection task. Activation of the intraparietal cortex was increased when problems were numerically more complex, while activation in the left inferior parietal cortex, particularly in the left angular gyrus was more related to the retrieval of well-learned problems and of rule-based information integration. Finally, activation of the right ventrolateral prefrontal cortex was associated with cognitive set changes complying with strategy use in the NBT.

We wish to emphasize that the differential activations reported in this study cannot rely on different task properties since the same task was used in all conditions. Fine-grain within-task manipulations have been previously used to examine and validate the neural correlates of one specific number representation (Naccache & Dehaene, 2001; Pinel et al., 2001). In the present study, different numerical representations were examined within one task and we suggest that such a within-task approach may be a promising tool to understand the interplay of different representations in the number processing network.

Study 6

Impairments of the mental number line for two-digit numbers in neglect

ABSTRACT

Humans represent numbers along a left-to-right orientated mental number line (MNL). Neglect patients seem to neglect the left part of the MNL, namely the smaller numbers within a given numerical interval. However, until now all studies examining numerical representation have focussed on single-digit numbers or two-digit numbers smaller than 50. In this study, the full range of two-digit numbers was assessed in neglect patients and two control groups. Participants were presented with number triplets (e.g. 10_13_18) and asked whether or not the central number is also the arithmetical middle of the interval. The factors manipulated were decade-crossing (e.g. 22_25_28 vs. 25_28_31), distance to the arithmetical middle (e.g. 18_19_32 vs. 18_24_32), and, most importantly, whether the central number was smaller or larger than the arithmetical middle (e.g. 11_12_19 vs. 11_18_19). Neglect patients differed from controls in that they benefited less when the middle number was smaller than the arithmetical middle of the interval. Neglect patients thus seem to have particular problems when accessing the left side of numerical intervals, also when adjusted to two-digit numbers. Such an impaired magnitude representation in neglect seems to have detrimental effects on two-digit number processing as the helpful spatial metric of magnitude cannot be properly activated.

INTRODUCTION

Patients suffering from neglect usually fail to attend to or respond to objects and/or people in the hemispace contralateral to their lesion site (Kerkhoff, 2000; Robertson & Halligan, 1999). When asked to bisect a physical line, neglect patients misallocate the true midpoint of the line and place it towards the right (Marshall & Halligan, 1989). Furthermore, neglect not only impairs the perception of visual objects but also affects representational space in general (Bisiach & Luzzatti, 1978).

While such symptoms of neglect concerning external space have been studied extensively, the impact of neglect on the representation of number magnitude as well as its processing has only more recently been focussed on (Goebel, Calabria, Farnè, & Rossetti, 2006; Priftis, Zorzi, Meneghello, Marenzi, & Umilà, 2006; Rossetti et al., 2004; Vuilleumier, Ortigue, & Brugger, 2004 and Zorzi, Priftis, & Umiltà, 2002; Zorzi, Priftis, Meneghello, Marenzi, & Umiltà, 2006). Number magnitude is assumed to be represented along a continuous (Priftis et al., 2006; Rossetti et al., 2004) and spatially organized *Mental Number Line* (MNL; Dehaene & Cohen, 1995; Restle, 1970), upon which smaller numbers are represented on the left and larger numbers are associated with the right (Dehaene, Bossini, & Gireaux, 1993; Gevers, Verguts, Reynvoet, Caessens, & Fias, 2006; Zorzi et al., 2006; Zorzi et al., 2002; Priftis et al., this issue, but see Vuilleumier et al., 2004).

In a seminal study, Zorzi et al. (2002) orally presented neglect patients with two numbers (e.g. “one” and “nine”) that defined a numerical interval and asked them to bisect this interval by naming the arithmetical middle (e.g. “five”) without doing any calculations. The size of the numerical interval was varied much as the line length is in the commonly used line bisection task. The authors observed that error rates increased as a function of the interval size as did the relative magnitude of the erroneous responses. For instance, all neglect patients produced errors such as falsely naming “6”, “7”, or “8” as being the midway between 1 and 9. On the other hand errors such as “2”, “3”, or “4” did not occur. So, neglect patients systematically neglected the left part of a numerical interval when bisecting it. This shifting to the *right* of a numerical interval closely mirrors the difficulties they commonly exhibit when bisecting a physical line.

However, until now almost all studies that have investigated number processing in neglect have focused on single-digit numbers (but see Rossetti et al., 2004). Moreover, all previous studies of neglect and number representation used the production version of the Number Bisection Task (NBT; e.g. Priftis et al., 2006; Rossetti et al., 2004 and Zorzi et al., 2002; 2006).

As pointed out by Nuerk, Geppert, van Herten, and Willmes (2002) a production version of the NBT may not be well-suited for assessing two-digit number processing due to the increased perceptual and cognitive demands imposed by two-digit numbers. Two-digit number processing requires the integration of different number representations (Nuerk et al., 2002): (i) spatial and magnitude information of a number complying with the place x value structure of the Arabic system (for a review, see Nuerk & Willmes, 2005), as well as (ii) a number being part of a multiplication table or not (i.e. multiplicativity; Nuerk et al., 2002).

However, Nuerk et al. (2002) showed that such different number representations can be assessed by a verification version of the NBT. In this verification version participants have to decide whether the central number of a triplet is also the arithmetical middle between the two outer numbers (i.e. a bisected triplet, e.g. 22_25_28) or not (i.e. a non-bisected triplet, e.g. 22_27_28).

In particular, Nuerk et al. (2002) observed that, among bisected triplets, multiplicatively related triplets (e.g. 21_24_27 vs. 22_25_28) were responded to faster and more accurately than non-multiplicative triplets (e.g. 25_28_31 vs. 22_25_28). As multiplication fact knowledge can be recruited to solve the task, more time-consuming magnitude manipulations can be bypassed (Delazer et al., 2006; Nuerk et al., 2002). Additionally, as known from other calculation tasks (e.g. carry-over effect in addition, Deschuyteneer, De Rammelaere, & Fias, 2005), triplets in which the interval crossed a decade boundary were more difficult than triplets in which the interval did not cross into the next decade (e.g. 22_25_28 vs. 25_28_31, see Nuerk et al., 2002).

For non-bisected triplets it was easier to classify triplets with a far distance of the (presented) central number to the actual arithmetical middle than triplets with a central number close to the arithmetical middle (e.g. 21_22_29 vs. 21_24_29; arithmetical middle 25). Most importantly, for the present study, an effect of the relative size of the central number was found (Nuerk et al., 2002; Moeller, 2006). Triplets with a central number smaller than the arithmetical middle of the interval (i.e. a relatively small central number within the interval, e.g. 18_19_32) are easier to reject than triplets with a central number larger than the arithmetical middle of the interval (e.g. 18_31_32).

But how do these item characteristics affect neglect patients? As elaborated above, neglect patients neglect the left side of a presented numerical interval. They do so in a production task and in particular for one-digit numbers. We suggest that this effect of neglect on the number magnitude representation generalizes to the full range of two-digit numbers and to the verification version of the NBT. Therefore, for non-bisected triplets we hypothesize that, unlike in healthy participants, the advantage for triplets involving central numbers smaller than the arithmetical middle of the numerical interval should be less pronounced or even disappear in neglect patients compared to non-neglect controls. If an interaction with distance to the middle were observed, the effect of size relative to the arithmetical middle should be more prominent for triplets with far distances to the middle and less prominent for close distances to the middle. This should be observed because, for far distances, the central number is much farther into the neglected part of the mental numerical interval.

For bisected triplets we do not expect strong modulations of the multiplicativity effect as this effect is not related to the spatial representation of number magnitude and was observed even in the absence of any distance effects in a patient with posterior cortical atrophy (Delazer et al., 2006). However, we hypothesize that the effect of decade crossing might be altered by neglect. Neglect patients may have particular difficulties with triplets in which a decade boundary is crossed: in such decade-crossing triplets, not only perception and representation of the units is required, but in addition the decade digits, which are perceptually and perhaps mentally on the neglected left side within a two-digit number, need to be processed. This can be accounted for by a deeper magnitude and base-ten processing required to correctly classify these triplets (Nuerk et al., 2002; Wood, Nuerk, & Willmes, 2006). More specifically, the impact of decade crossing should be more pronounced for those triplets that involve a decade crossing between the first and the central number as deeper magnitude evaluation on the left of a given numerical interval is needed.

In summary, neglect patients should be particularly affected (i) in triplets involving central numbers smaller than the arithmetical middle of the interval (which are presumably located on the left of the mental numerical interval) and (ii) in triplets crossing a decade boundary; they should not be affected by (iii) general distance properties or (iv) multiplicativity.

MATERIALS AND METHODS

Participants: Eighteen German-speaking participants took part in this study. Participants were pooled into three groups of 6 persons matched for age, gender, education and time post-lesion (4 males/2 females each, see Table 1): a patient group including 6 patients with right-sided lesions who suffered from left-sided neglect; a patient control group involving 6 patients with right-sided lesions but no clinical signs of neglect; and a healthy control group of 6 participants with no history of neurological or psychological illness. In all cases of stroke, (patient group and patient control group) lesions were confirmed by either MRI or CT. However, one patient (patient control group) was tested four years post-lesion and could not be matched for the parameter time post-lesion.

Visual field assessment showed no sign of hemianopia in any one of the participants. All participants had normal or corrected-to-normal vision, and were right-handed. Participation in this study was voluntary.

To diagnose neglect (patient group) and to rule out neglect (patient control group), a standardised neuropsychological neglect test-battery was carried out: Neglect Test (NET; Fels & Geissner, 1997 adapted German version of the Behavioural Inattention Test - BIT; Wilson, Cockburn, & Halligan, 1987). This test-battery includes 17 different tasks that can be allocated into the categories *conventional sub-tests* (e.g. line bisection, line and star crossing, figure and shape copying, representational drawing) and *behavioural subtests* (e.g. picture scanning, menu reading, article reading, telling the time from analogue and digital clock faces, set the time). Moreover, to rule out degenerative cognitive processing, the SIDAM (Structured Interview for the Diagnosis of dementia of the Alzheimer type, Multi-infarct dementia and dementias of other aetiology; Zaudig et al., 1996) was carried out. The SIDAM is a questionnaire for diagnosing dementia according to international diagnostic guidelines (ICD 10, DSM IV). This instrument includes simple questions and problems covering areas such as orientation, instantaneous recall, memory (short-, long-term), intellectual/cognitive abilities, verbal and numerical abilities, visual-spatial abilities, Aphasia and Apraxia. Additionally, the SIDAM also includes the Mini-Mental State Examination (MMSE; Folstein, Folstein, & McHugh, 2000), a dementia screening.

The participants' numerical and mathematical abilities were further evaluated using the EC 301 R (Claros Salinas, 1994). The EC 301 R is a test battery for assessing calculation and number processing in brain-damaged patients. It comprises the subtests: dot counting, free backward counting, number transcoding, mental arithmetic, array on a physical number line, number comparison (auditory as well as symbolically), multi-digit arithmetic (i.e. addition, subtraction and multiplication), as well as perceptive and contextual estimations.

None of the participants showed clinical signs of dementia or degenerative processes and all had good to perfect numerical and mathematical abilities, so that the EC 301 R showed no sign of acalculia in any one patient. Table 2 summarizes the test results of the NET, EC 301 R and SIDAM. For detailed results of the subtests of the NET (e.g. line crossing: patient M.M. correctly marked 17 out of 36 lines) and EC 301 R (e.g. dot counting: patient M.M. was not able to count the presented dots) see Appendix A.

Stimuli and design: 160 triplets of two-digit numbers (ranging from 11 to 99) were used in this study. The triplets were presented in Arabic notation. Overall distance, problem size, average parity, parity homogeneity, decade-crossing, the inclusion of decade numbers as well as tie numbers were matched between the respective stimuli groups.

80 non-bisected triplets were organized in a fully crossed 2 x 2 design incorporating the factors distance to the arithmetical middle (close vs. far) and size relative to the arithmetical middle (smaller or larger than the arithmetical middle). Distance to the middle refers to the distance of the presented middle number to the true arithmetical middle of the interval. In the triplet 21_22_35, the presented central number 22 has a far distance to the arithmetical middle (i.e. 28 - 22 = 6) while in 21_27_35, 27 is close to the arithmetical middle (i.e. 28 - 27 = 1). Size relative to the arithmetical middle refers to the position of the central number in relation to the arithmetical middle of the given interval. For the triplet 33_34_47 the presented middle number 34 is smaller than the arithmetical middle 40, while in 21_34_35 the number 34 is larger than the arithmetical middle 28.

Table 1: Demographic and clinical data of all participants

Neglect patients	**M.M.**	**K.R.**	**S.I.**	**S.G.**	**H.E.**	**R.R.**	*Mean(SD)*
Sex	Male	Male	Female	Male	Female	Male	
Age	57	46	45	62	55	63	*54.7(7.7)*
Education	8	13	9	8	9	9	*9.3(1.9)*
Time post lesion (weeks)	12	3	6	17	14	10	*10(5.2) Median 12*
Lesion etiology	IS	IS	HS(SAH)	HS	IS	IS	
Lesion site	RH; PO	RH;TP	RH;	RH; BN	RH; FP	RH; TP	
Affected blood vessel	Middle cerebral artery	Middle cerebral artery	Middle cerebral artery	Middle cerebral artery	Middle cerebral artery	Middle cerebral artery	
Control patients	**H.G.**	**F.C.**	**B.R.**	**B.H.**	**O.H.**	**G.D.**	
Sex	Male	Male	Female	Male	Female	Male	
Age	58	46	46	60	54	69	*55.5(8.8)*
Education	9	9	10	9	10	9	*9.3(0.5)*
Time post lesion (weeks)	10	6	5	44	17	200	*47(76.3)Median 14*
Lesion etiology	IS	IS	IS	IS	SAH	IS	
Lesion site	LH; P	RH; BS	RH; T	LH; Pons	Bilat. F	RH;FP	
Affected blood vessel	Middle cerebral artery	Artery verti-bralis	Cerebral posterior artery	Arteria pontis		Middle cerebral artery	
Healthy controls	**R.M.**	**Z.R.**	**L.H.**	**F.K.**	**S.G.**	**B.H.**	
Sex	Male	Male	Female	Male	Female	Male	
Age	57	47	48	63	55	67	*56.2(7.3)*
Education	12	9	8	9	9	13	*10(2)*

[1]This patient was tested four years post lesion and could not be matched for the parameter time post lesion

HS - hemorrhagic stroke, IS - ischemic stroke, SAH - subarachnoidal hemorrhagic

LH - left hemisphere, RH - right hemisphere

BN - basal nuclei, BS - brain stem, F - frontal, O - occipital, P - parietal, T - temporal

Table 2: Scores of each participant in the NET, SIDAM and EC 301 R; Individual mean RT and error rate

Neglect Group		**M.M.**	**K.R.**	**S.I.**	**S.G.**	**H.E.**	**R.R.**
NET	Standard value	78	163	136	99.5	147	71
	Diagnosis	Considerable	Low	Considerable - low	Considerable	Low	Very Considerable
SIDAM	Raw data	43/55[1]	54/55	50/55	37/55[1]	44/55[1]	52/55
EC301R	Raw data	84/135[2]	130/135	115/135[2]	Missing data.[3]	123/135[2]	103/135[2]
NBT (RT)	ms	8493	2425	7133	6436	5660	7452
NBT (Errors)	%	15.00	5.60	7.50	21.30	26.90	10.00
Patient group		**H.G.**	**F.C.**	**B.R.**	**B.H.**	**O.H.**	**G.D.**
NET	Diagnosis	No Negelct	No Negelct	No Negelct	No Negelct	No Negelct	No Negelct
SIDAM	Raw data	54/55	53/55	55/55	55/55	46/55	50/55
EC301R	Raw data	135/135	134/135	134/135	135/135	133/135	123/135
NBT (RT)	ms	2190	3332	2954	2738	3467	4421
NBT (Errors)	%	1.30	1.90	1.90	5.60	10.60	4.40
Healthy group		**R.M.**	**Z.R.**	**L.H.**	**F.K.**	**S.G.**	**B.H.**
NET	Diagnosis	No Negelct	No Negelct	No Negelct	No Negelct	No Negelct	No Negelct
SIDAM	Raw data	55/55	51/55	49/55	46/55	54/55	54/55
EC301R	Raw data	133/135	134/135	130/135	133/135	135/135	135/135
NBT (RT)	ms	3778	3336	5800	2860	5061	6387
NBT (Errors)	%	2.50	11.30	6.30	1.30	1.30	3.80

[1]Due to neglect symptoms some items of the SIDAM could not be successfully processed (e.g. copying shapes)
[2]Due to neglect symptoms some items of the EC310R could not be successfully processed (e.g. counting dots)
[3]This patient was tested on the following numerical and mathematical abilities: comparing single- and multi-digit Arabic numbers, mental arithmetic (addition, subtraction, multiplication, division) and written arithmetic (addition and subtraction). The patient's tested abilities were intact.

For 80 bisected triplets the factors multiplicativity (yes vs. no) and decade crossing (yes vs. no) were manipulated. Bisected triplets could not be fully crossed, as triplets which are part of a multiplication table larger than three would always cross into the next decade. Therefore, the factors decade-crossing (yes vs. no) and multiplicativity (yes vs. no) were assessed in distinct item sets.

The three numbers were presented above each other in the centre of the screen's right half. Each two-digit number extended 7 cm horizontally and 4.5 cm vertically. The numbers were vertically separated by 4.5 cm. The stimuli were presented in bold Arial font (size 48), white on black background.

In this study neglect patients' performance in the verification version of the NBT was compared to that of non-neglect controls.

Procedure: The experiment was run on a 1.6 GHz laptop and participants were seated approximately 50 cm in front of a 15´´ screen driven at a resolution of 1024x768 pixels.

The left and right arrow keys located at the bottom right-hand corner of the keyboard functioned as response buttons: bright orange stickers with the letters "J" for "yes" (in German "Ja") and "N" for "no" (in German "Nein") covered the left and right arrow keys, respectively. All other keys were covered up with white cardboard. Participants were instructed to indicate their decision by pressing one of the two response buttons as fast and as accurately as possible.

The instruction was followed by 20 practice trials incorporating one- and two-digit numbers. To ensure that patients perceived the presented numbers correctly all participants were asked to read out the numbers of the first four practice trials aloud. All patients were able to do so without mistakes. Participation in the critical experiment was only allowed when more than 2/3 (14 out of 20) practice trials were classified correctly. The experiment involved

4 blocks of 40 trials each, which the participant initiated by pressing one of the response keys. Trial order was randomized for each participant.

Before each trial a fixation cross was presented in the centre of the screen for 500ms. Then the three numbers were presented for 15,000 ms followed by a ISI of 5,000 ms. The experiment took approximately 15 to 25 minutes, depending on how long participants rested between blocks.

RESULTS

For statistical analyses RT and error data were used. Only RTs followed by a correct response were considered for further analysis. A trimming procedure eliminated trials with RTs shorter than 200 ms and longer than 15,000 ms and – in a second step – RTs outside +/- 3 SD of each participant's individual mean. In total 92.33 % of all trials fulfilled these criteria and were used for further analyses. As stroke patients show a greater variability in their RTs than healthy participants, a *z*-transformation on individual item RT, using the mean and standard deviation per participant for standardization, was carried out. Error rates were arcsine transformed prior to the analyses. There was no speed accuracy trade-off present in our data as indicated by the absence of a reliable negative correlation of error rate and RT in any item condition (all $r > -.40$, all $p > .10$, $N = 18$).

The main interest of the current study was to contrast patients suffering from neglect and those without neglect in terms of their performance in the NBT. Therefore, the two control groups (patient controls and healthy controls) were pooled into one group for the main analyses of this study. For the interested reader Appendix B reports all results separated for the single control groups. As neglect patients do not only neglect the left side of space but also the left parts of numerical intervals (e.g. Zorzi et al., 2002), processing numbers on the left side of a numerical interval (i.e. the central number being smaller than the arithmetical middle of the interval) should be relatively more impaired.

Factorial Analyses (ANOVA and t-tests)

Non-bisected triplets (RT and error data):

Response latencies: An analysis of variance (ANOVA) incorporating the factors distance to the middle, size relative to the middle and group was conducted on z- transformed RTs. Most importantly, and in accordance with the hypothesis, a reliable interaction between size relative to the arithmetical middle and patient group was obtained [see Figure 1A, $F(1, 16) = 6.40$, $p < .05$]. This indicates that neglect patients did benefit reliably less from a central number smaller than the triplet's arithmetical middle than did control participants (426 ms vs. 637 ms). However, this disadvantage for neglect patients did not seem to be modulated by distance to the arithmetical middle. This was indicated by the non-significant three-way interaction of distance to the middle, interval side and group [$F(1, 16) < 1$]. However, from the notion of a spatially oriented MNL (e.g. Dehaene et al., 1993), it can be inferred that neglect patients should exhibit a more pronounced disadvantage in triplets such as 21_22_35 compared to e.g. 21_27_35. On the one hand, both central numbers are smaller than the arithmetical middle of the interval (e.g. 28) in these examples. But on the other hand, they differ in their distance to the middle: while 27 is numerically close to the arithmetical middle of the interval (e.g. 28), 22 is far smaller than the middle. Assuming an ascending left-to-right orientation of the MNL, 22 would be located further to the left than 27 and should thus be even more neglected. To directly test this hypothesis the absolute effects of the factor size relative to the arithmetical middle were contrasted for neglect patients vs. non-neglect controls separately for triplets with a central number either numerically close to the arithmetical middle or far from it using Bonferroni-Holm corrected *t*-tests (Holm, 1979). The beneficial effect of the central number being smaller than the arithmetical middle only tended to be less pronounced for neglect patients [$t(16) = 1.95$, $p = .07$, one-sided], whereas neglect patients benefited reliably less from a second number far smaller than the arithmetical middle [$t(16) = 2.19$, $p < .05$, one-sided]. This indeed indicates a greater relative disadvantage for neglect patients for triplets involving a central number far smaller than (i.e. further to the left of) the arithmetical middle.

Additionally, the ANOVA showed strong main effects of distance to the arithmetical middle [$F(1, 16) = 47.44$, $p < .001$] and size relative to the middle [$F(1, 16) = 27.18$, $p < .001$]. So, triplets in which the central number was numerically far from the arithmetical middle were rejected faster than triplets with a central number numerically close to the middle (4759 ms vs. 5470 ms). Moreover, triplets with a second number smaller than the arithmetical middle were responded to faster than triplets with a second number larger than the middle (4849 ms vs. 5380 ms). Furthermore, the interaction of distance to the middle and size relative to the middle was reliable [$F(1, 16) = 15.02$, $p < .01$]. This means that for far numerical distances to the middle the beneficial effect of a central number smaller than the arithmetical middle was more pronounced than for close distances (948 ms vs. 114 ms). Finally, the two groups, i.e. neglect and no neglect did not differ reliably in terms of RT [$F(1, 16) < 1$].

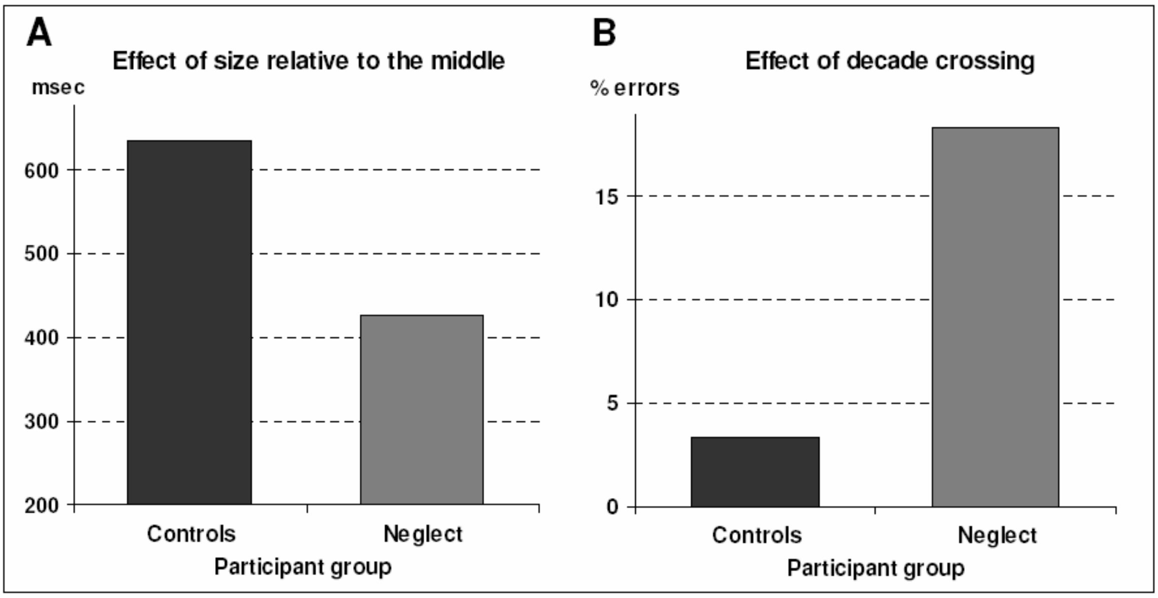


Figure 1: *The beneficial effect in ms for triplets with a central number smaller than the arithmetical middle of the interval separately for Non-neglect controls and neglect patients is depicted in Panel A. Panel B gives the detrimental effect of decade crossing in % errors separated for non-neglect controls and neglect patients.*

Errors: The ANOVA revealed a main effect of participant group [$F(1,16) = 13.52$, $p < .01$], indicating that neglect patients committed reliably more errors than non-neglect controls (7.9 % vs. 2.9 % errors). Additionally, the interaction of distance to the middle and size relative to the middle was significant [$F(1, 16) = 4.84$, $p < 05.$]. This indicates that, for triplets with a central number numerically close to the arithmetical middle, rejection was more error-prone when the central number was smaller than the middle compared to when it was larger than the middle (7.9 % vs. 4.8 % errors, respectively). However, this pattern was reversed for triplets with a central number far from the middle (smaller: 2.5 % vs. larger: 6.5 %). There were no further significant main effects for the manipulated item characteristics [all $F < 2.43$, all $p > .14$] or interactions of group with any of them [all $F < 2.35$, all $p > .15$].

Bisected triplets (RT and error data):

Decade crossing: In line with the hypothesis, a marginally significant interaction of decade crossing and participant group was observed for errors [$F(1, 16) = 4.06$, $p = .06$], but not for response latencies [$F(1, 16) < 1$]. When directly testing the hypothesis of neglect patients' performance being more impaired for triplets crossing into the next decade, the absolute effects of decade crossing were contrasted for neglect patients vs. non-neglect controls. The *t*-test revealed [$t(16) = 2.02$, $p < .05$, one-sided] that the increase of errors in triplets crossing a decade boundary was more pronounced in neglect patients (+ 18.4 % errors, see Figure 1B) than in non-neglect controls (+ 3.3 % errors). However, one could argue that neglect patients committed more errors as they were not able to correctly perceive and identify the decade digits of the two-digit numbers constituting the numerical interval. To evaluate this potential constraint a regression analysis on item RTs was conducted incorporating problem size (operationalised by the arithmetic mean of a triplet) as predictor[1]. It was observed that problem size reliably predicted item RT [$b = .21$, $t(159) = 2.67$, $p < .01$]

[1] Please note that there were too many perfect scores to run a meaningful analysis on mean item error rates.

meaning that larger problem size was associated with longer response latencies. This suggests that neglect patients were not only able to perceive the decade digits of the involved two-digit numbers but also processed their magnitude.

Furthermore, a strong main effect for decade-crossing was found for both speed and accuracy [RT: $F(1, 16) = 34.59$, $p < .001$, accuracy: $F(1, 16) = 15.95$, $p < .001$]. Triplets that did not cross a decade boundary were responded to faster (4620 ms) and more accurately (5.4 % errors) than triplets which crossed into the next decade (5195 ms and 16.3 % errors). Exclusively for response latencies this effect was even further differentiated: while non-neglect controls benefited (-168 ms) from a decade crossing occurring between the first and the central number of a triplet (e.g. 28_31_34 vs. 25_28_31), this had a deteriorating effect on neglect patients [+908 ms, $t(16) = 2.79$, $p < .05$]. This suggests that the cognitively more demanding processing of a decade crossing is especially impaired when it occurs in the neglected part of the numerical interval.

Moreover, the ANOVA revealed a reliable difference between neglect patients and non-neglect controls for RT [$F(1, 16) = 5.85$, $p < .05$] as well as errors committed [$F(1, 16) = 8.06$, $p < .05$]. So, neglect patients exhibited longer response latencies (6330 ms) and made more errors (17.4 %) compared to non-neglect controls (3485 ms and 4.2 % errors).

Multiplicativity: A main effect for multiplicativity was observed neither for RT nor for errors [RT: $F(1, 16) < 1$, errors: $F(1, 16) = 2.02$, $p = .17$]. However, neglect patients and non-neglect controls differed significantly in respect to errors committed [$F\ (1, 16) = 11.23$, $p < .01$] but not for response latencies [$F\ (1, 16) = 1.09$, $p = 31$]. Neglect patients' performance was more error prone (24.2 % errors) than that of non-neglect controls (7.3 % errors). Finally, multiplicativity and participant group did not interact reliably for either RT or errors [RT: $F(1, 16) < 1$, errors: $F(1, 16) < 1$].

DISCUSSION

The effects of neglect on the bisection of numerical intervals in the present study can be summarized as follows: First, the ANOVA revealed that neglect patients benefited reliably less when the central number was smaller than the arithmetical middle (e.g. 21_22_35) compared to when it was larger (e.g. 21_34_35). Second, it was observed that neglect patients were especially impaired for triplets crossing a decade boundary (e.g. 25_28_31) compared to triplets within the same decade (e.g. 22_25_28). Finally, neglect patients were not generally impaired in all numerical representations: Neither the effect of distance to the arithmetical middle nor that of multiplicativity differed between neglect patients and non-neglect controls.

According to Nuerk et al. (2002; see also Moeller, 2006), participants' response latencies are shorter for triplets with the central number being numerically smaller than the arithmetical middle compared to triplets with a central number larger than arithmetical middle. This general beneficial effect was replicated in the present study. However, neglect patients benefited reliably less than non-neglect controls when the central number was smaller than the arithmetical middle. From this it can be inferred that neglect patients neglect the left side of a given numerical interval.

It can be hypothesized that the effect of neglect on the mental numerical interval is stronger for numbers farther towards the left side of that interval than for numbers which are on the left side of the interval but relatively close to the middle. Although there was no reliable three-way interaction between group, size relative to the middle, and distance to the middle, fine-grained analyses revealed that this hypothesis tended to be true. While for numbers close to the arithmetical middle, neglect patients and controls differed only marginally; for numbers far from the middle, there was a significant group effect: For those far distances, neglect patients benefited significantly less when the central number was smaller than the arithmetical middle than when it was larger. These findings extend previous results from Zorzi et al. (2002; and later Priftis et al., 2006; Rossetti et al., 2004 and Zorzi et al. 2006). Neglect not only impairs access to relatively smaller numbers in the one-digit or small two-digit range, it does so for the full range of two-digit numbers.

The second hypothesis outlined in the introduction was that neglect patients may have particular difficulties with two-digit number triplets which cross a decade boundary compared to triplets from within one decade. Indeed, we did not only observe a general decade crossing main effect, but a reliable interaction with participant group: Neglect patients exhibited a larger decade crossing effect than their non-neglect counterparts; i.e. they had particular difficulties when the numbers of a triplet were from different decades.

The effect of decade crossing on behavioural performance (Nuerk et al. 2002) and fMRI signal (Wood et al., 2008) has been interpreted as resulting from a deeper processing of number magnitude and the base-10 structure of the Arabic number system required when different decade digits are encountered in the task. Processing different decades seems to demand the integration of decomposed decade digit and unit digit magnitudes (cf. Nuerk & Willmes, 2005 for a discussion). Neglect patients may have problems mentally representing the magnitude of two-digit numbers in a decomposed fashion, as they may have problems representing the decade digit's magnitude. This may lead to the more pronounced decade crossing effect observed for neglect patients in the current study.

A perceptual account for this larger decade crossing effect is that neglect patients may have difficulties *perceiving* the decade digit which is located on the left within a two-digit number. This perceptual account is however not consistent with the diagnostic data of our patients: All neglect patients were able to read aloud the two-digit numbers of a triplet presented above each other without errors. Moreover, in the NET, the neglect patients were even able to read four-digit digital clock times without major problems (one single error in two of six patients). Moreover, this perceptual account is also disproved by more fine-grained analyses: (Decade digit driven) problem size was a highly significant predictor of neglect patients' performance. So, neglect patients did not only perceive the visual presence of the decade digits but they even processed their magnitude. Finally, we wish to point out that the visual presentation of the number triplets with the numbers above each other was identical in all experimental conditions. Only the mental spatial representation in the mental numerical interval differed between conditions.

For all these reasons, it can be suggested that the differences between conditions are not due to perceptual impairments in neglect, but to impairments of the mental (spatial) representation of the presented two-digit numbers.

Interestingly, the group difference in the decade crossing effect seemed to be modulated by the location of this decade crossing within the numerical interval: A detailed comparison of triplets with decade crossing occurring between the first and the central number (e.g. 28_31_34) and triplets with decade crossing between the central and the third number (e.g. 25_28_31) revealed that the deficit present for neglect patients was more pronounced in triplets with decade crossing between the first and the central number. In those triplets, the decade crossing is on the left side of the mental numerical interval to which access seems particularly impaired in neglect. Therefore, this detailed analysis of the decade crossing effect corroborates the assumption of an oriented MNL influencing performance in the NBT.

In the remaining paragraphs, the results regarding hypotheses (iii) and (iv) will be discussed: With regard to the factors distance to the arithmetical middle and multiplicativity no group differences between neglect patients and non-neglect controls were observed.

As the distance to the arithmetical middle effect is a magnitude-related effect, it may seem surprising at first that no group difference was observed because, for the other magnitude-related factors, neglect patients showed specific impairments as discussed above. For the distance to the middle effect, this was different. A possible reason for this lack of difference between the two groups may be that magnitude processing can be subserved by the IPS bilaterally (cf. Dehaene et al., 2003). General difficulties to process number magnitude (e.g. with regard to the distance effect) are usually only observed when both IPS areas are affected (e.g. Dehaene & Cohen, 1997; Delazer et al., 2006, but see Ashkenazi, Henik, Ifergane, & Shelef, this issue for differing results). However, the left IPS was preserved in our neglect patients as they all had right-hemispheric lesions. Thus, this left IPS area might have compensated for the loss of the right IPS with regard to general magnitude and numerical distance processing.

Nuerk et al. (2005) employed a magnitude comparison task in an Eriksen flanker paradigm. They found that the magnitude of the distractors was processed. However, the magnitude of the distractors was not spatially represented on a MNL as no SNARC effect was observed for the distractors' magnitudes while a robust distance effect was present. Consequently, number magnitude can be processed without automatic activation of its spatial code. This spatial code may only be activated in conditions in which spatial processing is salient and does not require much attention.

In the current study, the mental numerical interval may be the most salient spatial reference frame. If the right IPS region were mainly responsible for linking number magnitude and spatial representation, damage to this right IPS region would be particularly impairing for this most salient spatial reference frame. In contrast, magnitude processing which is less spatially salient may be less impaired as the left IPS, which may be less involved in spatial representation of magnitude, can take over. Although this hypothesis of differential involvement of the two IPS in the spatial processing of magnitude offers an account for the result pattern observed in the current study, it should be further investigated in the future.

Finally, the fact that we did not observe any group differences with regard to multiplicativity may be due to two reasons: First, multiplication fact knowledge is presumably located in the left gyrus angularis (cf. Dehaene et al., 2003; Delazer et al., 2003) which was not lesioned in our neglect patients (see Table 1 for lesion sites of neglect patients). Second, in contrast to previous studies, we did not even observe a main effect of multiplicativity. The absence of an effect of multiplicativity may be explained by two related factors: first, the low strategic value of retrieving this information, since multiplicative triplets formed only 12.5 % of all triplets instead of 25 % in the original study by Nuerk et al. (2002); and second, by reduced executive resources of both controls and patients. In the original study by Nuerk et al. (2002), 25 % of all triplets were part of a multiplication table. Therefore, activating the representation of multiplicativity was relevant for improving performance, as half of the "Yes" responses could be determined by the recognition of multiplicatively related triplets. In the present study, only 12.5 % of all triplets were part of a multiplication table. This difference in the proportion of multiplicatively related triplets may have an impact on the strategic relevance of multiplicativity. As it is strategically less relevant in the present study, the effect of multiplicativity may have disappeared.

Moreover, automatically activating multiplication fact knowledge may have been hampered by general reductions in the capacity of neglect patients and control participants to monitor and integrate information from different sources in parallel due to age. In a recent fMRI study on the NBT, Wood et al. (submitted) showed that the left angular gyrus and in the superior frontal gyrus were sensitive to multiplicativity. While the left angular gyrus may be associated with the retrieval of multiplication facts (cf. Dehaene et al., 2003; Delazer et al., 2003), the superior frontal gyrus was linked to the monitoring of potentially relevant information from different sources (e.g. Brass & von Cramon, 2005). Although there is no reason to doubt that older patients as well as neglect patients can access the representation of multiplicativity, they may not have monitored for the multiplicativity in the present study due to limitations in executive resources and working memory (Doricchi, Guariglia, Gasparini, Tomaiuolo, 2005).

CONCLUSIONS

Neglect impairs number processing of two-digit numbers in a specific way: neglect patients seem to neglect the left part of a given numerical interval activated on a spatially oriented MNL. Moreover, our data also indicate that neglect patients have specific difficulties in processing decomposed magnitudes of two-digit numbers. In particular, neglect patients may have difficulties accessing the magnitude of the decade digit, which is on the left within a two-digit number. However, diagnostic data and additional analyses suggested that this deficit was not due to perceptual limitations.

These results extend previous findings in several theoretical and methodological aspects. Neglect not only impairs the spatial numerical representation of one-digit and small two-digit numbers, but also the representation of the whole two-digit number range. In particular, neglect not only impairs the spatial holistic magnitude representation of two-digit numbers in general, it specifically impairs the integration of tens' and units' magnitude representation of two-digit numbers. Finally, neglect not only impairs performance in actively producing the middle number of a numerical interval, it also impairs the verification of whether a given number is the arithmetical middle of an interval or not.

APPENDIX A

Test results of patients suffering from neglect in all sub tests of the NET

NET	Patients					
Subtests (max. raw score)	**M.M.**	**K.R.**	**S.I.**	**S.G.**	**H.E.**	**R.R.**
Line crossing (36)	17	36	36	18	36	33
Letter cancellation (40)	26	37	33	26	34	13
Star crossing (54)	15	54	47	19	51	18
Figure copying (star) (3)	0	3	2	0	3	0
Figure copying (rhombus) (3)	2	3	0	3	3	1
Figure copying (flower) (3)	0	3	2	2	2	0
Line bisection (9)	2	9	5	6	5	2
Representational drawing (3)	2	2	3	2	2	0
Article reading (140)	66	140	93	93	93	71
Copy address (88)	50	88	85	70	85	61
Picture scanning (menu) (8)	6	8	8	6	8	5
Picture scanning (sink) (9)	9	8	9	6	9	2
Picture scanning (room) (15)	13	14	15	8	14	5
Menu reading (24)	19	24	24	12	12	11
Tell the time (digital) (3)	3	3	2	3	3	2
Tell the time (analogue) (3)	1	3	3	1	3	2
Set the time (analogue) (3)	2	3	2	1	3	3
Σ Raw score[1]	**233**	**438**	**369**	**270**	**366**	**229**
Σ Standardized scores[2]	**78**	**163**	**136**	**99**	**147**	**71**

[1]Σ max. raw score for all subtests 444
[2]Σ max. standardized scores for all subtests 170 (Diagnostic guidelines for neglect: 0-72 very considerable, 73-135 considerable, 136-166 low)

Test results of patients suffering from neglect in all sub tests of the EC 301 R

EC 301 R	Patients					
Subtests (max. raw score)	**M.M.**	**K.R.**	**S.I.**	**S.G.**[3]	**H.E.**	**R.R.**
Counting dots (6)	0[2]	5	6		6	3[2]
Free backward counting (2)	2	2	2		2	2
Transcoding						
→ writing (12)	12	12	12		12	11
→ reading (12)	12	12	12		12	12
→ 1 ⇨ one (12)	0[2]	12	12		0[2]	4[2]
Mental arithmetic (16)	15	16	15		16	16
Physical number line (10)	8	10	10		10	10
Number comparison						
→ auditory (16)	16	16	16		16	16
→ written (16)	0[2]	16	4[1]		16	4[1]
Multi-digit arithmetic						
→ addition (4)	0[2]	2	4		4	4
→ subtraction (4)	0[2]	4	4		4	4
→ multiplication (7)	3	7	2		7	1[2]
Perceptive estimations (8)	8	8	8		8	6
Contextual estimations (10)	8	8	8		10	10
Σ Raw score[1]	**84**	**130**	**115**		**123**	**103**

[1]Σ max. raw score for all subtests 135
[2]Due to neglect symptoms this subtest could not be carried out successfully
[3]This patient was tested on the following numerical and mathematical abilities: comparing single- and multi-digit Arabic numbers, mental arithmetic (addition, subtraction, multiplication, division) and written arithmetic (addition and subtraction). The patient's tested abilities were intact.

APPENDIX B

Factorial analyses (ANOVA and *t*-tests) differentiating between all three participant groups:

Non-bisected triplets (RT and error data):

Response latencies: An analysis of variance (ANOVA) incorporating the factors distance to the middle, size relative to the middle and participant group was conducted on z-transformed RTs. The ANOVA revealed that the interaction between size relative to the arithmetical middle and patient group was marginally significant [$F(2, 15) = 3.01$, $p = .08$]. Subsequently, Bonferroni-Holm corrected *t*-tests were conducted to directly test the hypothesis that neglect patients should benefit less from a central number smaller than the arithmetical middle of the interval: Neglect patients did benefit reliably less (-426 ms) from a second number smaller than the triplet's middle than did either patient controls [-527 ms, $t(10) = 1.82$, $p = .05$, one-sided] or healthy controls [-747 ms, $t(10) = 2.15$, $p < .05$, one-sided[a]]. The three-way interaction of distance to the middle, size relative to the middle and participant group [$F(2, 15) < 1$] was not reliable indicating that the disadvantage for neglect patients was not modulated by distance to the arithmetical middle.

Additionally, the ANOVA showed strong main effects for of distance to the middle [$F(1, 15) = 48.29$, $p < .001$] and size relative to the middle [$F(1, 15) = 38.71$, $p < .001$]. So, triplets in which the central number was numerically far from the arithmetical middle were rejected faster than triplets with a central number numerically close to the middle (4424 ms vs. 5043 ms). Moreover, triplets with a second number smaller than the arithmetical middle were responded to faster than triplets with a second number larger than the middle (4451 ms vs. 5017 ms). Furthermore, the interaction of distance to the middle and size relative to the middle was reliable [$F(1, 15) = 13.94$, $p < .01$]. This means that for far numerical distances to

[a] Note: Similar results were obtained when directly contrasting the effects of decade crossing in an one-way ANOVA with subsequent post-hoc Dunnett *t*-tests: neglect vs. patient control $p = .05$, neglect vs. healthy control $p < .05$.

the middle the beneficial effect of a central number smaller than the arithmetical middle was more pronounced than for close distances (930 ms vs. 203 ms). Finally, the three participant groups did not differ reliably in terms of RT [$F(2, 15) < 1$].

Errors: In the error analysis a main effect of group was observed [$F(2,15) = 6.83$, $p < .01$]. Bonferroni-Holm corrected *t*-tests showed that neglect patients committed reliably more errors than control patients [+5.0 %, $t(10) = 3.41$, $p < .01$, one-sided] as well as healthy controls [+5.0 %, $t(10) = 3.12$, $p < .01$, one-sided]. Additionally, the interaction of distance to the middle and size relative to the middle was significant [$F(1, 15) = 5.23$, $p < 05.$]. This indicates that for triplets with a central number numerically close to the arithmetical middle rejection was more error prone when than when the central number was smaller than the middle compared to when it was larger than the middle (6.4 % vs. 3.9 % errors, respectively). However, this pattern was reversed for triplets with a central number far from the middle (smaller: 1.9 % vs. larger: 6.1 %). There were no further significant main effects of the manipulated item characteristics [all F < 1.16, all p > .30 or interactions of group with any of them [all F < 1.11, all p > 13].

Bisected triplets (RT and error data):

Decade crossing: In line with the hypothesis, the interaction of decade crossing and participant group was reliable for errors [$F(2, 15) = 6.53$, $p < .01$], but not for response latencies [$F(2, 15) < 1$]. Bonferroni-Holm corrected *t*-tests were conducted to directly test the hypothesis of neglect patients' performance being impaired most for triplets crossing into the next decade. Therefore, the absolute effects of decade crossing were contrasted for neglect patients vs. patient controls [$t(10) = 0.90$, $p = .20$, one-sided] as well as for neglect patients vs. healthy controls [$t(10) = 2.82$, $p < .05$, one-sided[b]]. This indicates that the increase of errors

[b] Note: Again, similar results were obtained when directly contrasting the effects of decade crossing in an one-way ANOVA with subsequent post-hoc Dunnett *t*-tests: neglect vs. patient control $p = .37$, neglect vs. healthy control $p < .01$.

for triplets crossing a decade boundary was more pronounced in neglect patients (+ 18.4 % errors) than in healthy controls (- 0.8 % errors) but not when compared to patient controls (+ 7.5 % errors).

Furthermore, a strong main effect for decade-crossing was found for both speed and accuracy [RT: $F(1, 15) = 40.74$, $p < .001$, accuracy: $F(1, 15) = 17.38$, $p < .001$]. So, triplets that did not cross a decade boundary were responded to faster (4164 ms) and more accurate (4.4 % errors) than triplets which crossed into the next decade (4702 ms and 12.8 % errors). Moreover, the ANOVA revealed that the three participant groups differed reliably regarding errors committed [$F(2, 15) = 3.78$, $p < .05$] but not RT [$F(2, 15) = 2.76$, $p = .10$]. Bonferroni-Holm corrected *t*-tests confirmed the hypothesis that neglect patients did more errors compared to patient controls [+13.7 %, $t(10) = 2.22$, $p = .05$, one-sided] as well as healthy controls [+13.7 %, $t(10) = 1.94$, $p < .05$, one-sided].

<u>Multiplicativity:</u> A main effect for multiplicativity was observed neither for RT nor for errors [RT: $F(1, 15) < 1$, errors: $F(1, 15) = 1.59$, $p = .23$]. However, neglect patients and non-neglect controls differed significantly for errors committed [$F(2, 15) = 5.27$, $p < .05$] but not for response latencies [$F(2, 15) < 1$]. Again, Bonferroni-Holm corrected *t*-tests showed that neglect patients' performance was more error prone compared to that of patient controls [+16.7 %, $t(10) = 2.62$, $p < .05$, one-sided] as well as healthy controls [+14.1 %, $t(10) = 2.72$, $p < .05$, one-sided]. Finally, multiplicativity and participant group did not interact reliably for either RT or errors [RT: $F(2, 15) < 1$, errors: $F(2, 15) < 1$].

Section 5

A computational model of place-value integration in two-digit number comparison

<u>Study 7</u>

Two-digit number processing – holistic, decomposed or hybrid?

A computational modelling approach

ABSTRACT

Currently, there are three competing theoretical accounts of the nature two-digit number magnitude representation: a holistic, a strictly decomposed, and a hybrid model. Observation of the unit-decade compatibility effect (Nuerk et al., 2001) challenged the view of two-digit number magnitude to be represented as one integrated entity. However, at the moment there is no way to distinguish between the decomposed and the hybrid model empirically. The present study addressed this issue using a computational modelling approach. Three network models complying with the constraints of either of the three theoretical models were programmed and trained on two-digit number comparison. Evaluation of the data produced indicated that by means of their power to account for empirical effects in the most parsimonious way the empirical data were simulated best by the strictly decomposed model. Although the more complex hybrid model accounted for all empirical effects as well. Implications of these results on our understanding of the nature of the human number magnitude representation are discussed.

INTRODUCTION

The internal representation of numerical magnitude is a central component in all cognitive models of number processing in humans (e.g., Campbell, 1994; Cipolotti & Butterworth, 1995; Dehaene & Cohen, 1995, 1997; Dehaene, Piazza, Pinel, & Cohen, 2003; McCloskey, 1992; for an overview see Deloche & Willmes, 2000). These models differ in the number and kind of postulated representations and their interactions, but the goal of all attempts was to develop a model that is sufficiently detailed to explain magnitude processing in both normal and cognitively impaired persons. There are at least two empirical findings that any proposed form of number magnitude representation should account for: the distance effect and the problem size effect. First, the distance effect indicates increased speed and precision in discriminating between two numbers with increasing numerical distance between them (e.g., Hinrichs, Yurko, & Hu, 1981; Moyer & Landauer, 1967). Second, the problem size effect denotes that performance in number processing or calculation worsens with increasing numerical magnitude of the numbers involved (e.g., Brysbaert, 1995; see also Zbodroff & Logan, 2005 for a review).

However, the specifics of mental number magnitude representation, which can account for these basic effects in numerical cognition, are still under debate. On the one hand, most models of number processing agree about (or at least do not oppose) the existence of one (or several) mentally represented number line(s) involved in numerical tasks such as magnitude comparison or mental arithmetic. On the other hand, the way this number line is organized in detail as well as the characteristics driving the mapping of numbers onto this mental number line is subject to controversial discussions among researchers. To address this issue, the case of two-digit numbers will be dealt with in the following section.

Two-digit number processing

Basically, Arabic numbers are structured in a two-dimensional place-value system. First, the base dimension is reflected by the ten digits from 0 to 9, whereas, second, the power dimension is coded by the positions of the digits with a base ten. The position within a digit string defines the value of each digit on a power of ten dimension (cf. Zhang & Norman, 1995; Chrisomalis, 2004). However, the way in which these two dimensions of numbers are represented psychologically on the mental number line is subject to controversial debates. At the moment three models accounting for the organization of this number line can be discerned.

1. The holistic model

According to the holistic model (e.g., Dehaene, Dupoux, & Mehler, 1990) numbers are transformed into an integrated entity mapped onto a holistic magnitude representation before any magnitude comparison or numerical manipulation is performed. In the context of number magnitude comparison this property implies that performance should be exclusively determined by the overall distance between the numbers to-be-compared, thereby, overriding the base-10 structure of two-digit numbers.

However, to account for the problem size effect additional assumptions are required. In this context, either linear coding of numbers along the mental number line with scalar variability or logarithmic coding of numbers with fixed variability have been suggested (for a recent discussion see Brannon, Wusthoff, Gallistel, & Gibbon, 2001; Dehaene, 2001). The latter model implies that magnitudes of larger numbers may be compared more slowly because magnitude is assumed to be represented along a logarithmically compressed mental number line. Consequently, the distance between the logarithmic magnitudes of two relatively larger numbers is smaller than the distance between the logarithmic magnitudes of two relatively smaller numbers. As variability is assumed to be constant for the representation of all numbers the overlap for larger numbers is suggested to be gets relatively larger. Due to this larger overlap more interference with the magnitude representations of neighbouring numbers occurs slowing down magnitude comparison of larger numbers compared to magnitude comparison of two smaller numbers with the same absolute distance. Alternatively, when linear coding with scalar variability is assumed, numbers are mapped with equal separations onto the mental number line but the variability of their entries increases with their size. Again, the overlap between two larger numbers is relatively larger and the comparison is slowed down correspondingly.

2. The decomposed model

The decomposition model (e.g., Verguts & De Moor, 2005) predicts that the magnitudes in two-digit numbers are represented separately for each of the constituting digits. Thus, the magnitude of a two-digit number is not mapped onto one mental number line as an entity. Instead, each digit constituting a two-digit number is mapped onto its own number line. Considering the base-10 structure of the Arabic number system, these mental number lines must then be labelled by referring to the place-value of the digit represented. In number comparison, separate comparisons of corresponding digits at equivalent positions (e.g., units, tens, etc.) within the to-be-compared numbers are assumed. The representation of the individual numbers (i.e., 0 - 9) on the separate number lines is supposed to be organized in an analogue way comparable to the one postulated by the holistic model of magnitude representation. This assumption allows accounting for the problem size effect based on the increasing magnitude of the individual digits on their respective number lines (see above for the distinction between linear coding with scalar variability and logarithmic coding with fixed variability).

3. The hybrid model

The hybrid model (e.g., Nuerk, Weger, & Willmes, 2001; Nuerk & Willmes, 2005) suggests that both, decomposed as well as holistic representations of number magnitude get activated in a comparison task. Magnitude comparison is assumed to operate in parallel on both representations, but depending on task requirements either one may be of more relevance (e.g., the holistic representation for approximation). As the hybrid model is assumed to incorporate characteristics of both of its constituting models the distance as well as the problem size effect are accounted for accordingly.

Because each of the three models can account for the distance as well as the problem size effect, another distinguishing feature is required. Nuerk and colleagues presented evidence suggesting "decade breaks in the mental number line" (Nuerk et al., 2001, p. B25): the unit-decade compatibility effect. In the next paragraph a distinction between the three approaches based on the compatibility effect will be discussed, and open questions will be put forward afterwards.

The Compatibility effect

The decade-unit-compatibility effect (Nuerk et al., 2001; Nuerk, Weger, & Willmes, 2002; 2004a; 2005; Moeller, Fischer, Nuerk, & Willmes, 2009a; Pixner, Moeller, Zuber, & Nuerk, 2009; Wood, Nuerk, & Willmes, 2006) describes the finding that it takes participants significantly longer to single out the larger number of e.g. the pair 38_53 than the pair 42_57. This is supposed to be the case because in the first example separate comparisons of tens and units lead to transient incompatible decision biases (i.e., 38_53 → $3 < 5$, but $8 > 3$) whereas in the latter, compatible pair of numbers no such incompatibility occurs (i.e., 42_57 → $4 < 5$ and $2 < 7$). Because the absolute overall distance is 15 in both examples, no compatibility effect should be observed if an exclusively analogue (holistic) magnitude representation for two-digit numbers were engaged. Additionally, the observed interaction of compatibility with unit distance (i.e., a more pronounced compatibility effect for large as compared to small unit distances) indicates that no common attentional congruity effect can be taken to be responsible for that finding. Thus, a model accounting for the compatibility effect must allow to differentiate between processing the tens numeral and the unit numeral – hence a model of decomposed two-digit number processing.

Nuerk and co-workers repeatedly observed (logarithmic) overall distance to be a reliable predictor of item RT in multiple stepwise regression analyses (cf. Nuerk et al., 2001; 2002; Moeller et al., 2009a; Wood et al., 2006). The authors interpreted the impact of overall distance between the to-be-compared numbers to be due to a holistic representation of the overall magnitude of the numbers involved, which cannot be accounted for by strictly decomposed representations of tens and units (see Nuerk & Willmes, 2005, for a detailed discussion of this argument). Based on these considerations, Nuerk and Willmes (2005) proposed the hybrid model of two-digit number processing (see also Nuerk et al., 2001). The decomposition part of the model accounts for the compatibility effect (which is hard to reconcile with a purely holistic magnitude representation), while the holistic part of the model would account for the influence of (logarithmic) overall distance.

Taken together, there is considerable empirical evidence for the existence of decomposed processing in multi-digit numbers (e.g., Nuerk et al., 2001; Moeller et al., 2009a; see also Domahs, Delazer, & Nuerk, 2006; Hoeckner et al., 2008; Nuerk et al., 2002; Ratinckx, Nuerk, van Dijk, & Willmes, 2006; Verguts & de Moor, 2005; for further evidence on decomposed processing other than the compatibility effect). So far there is no empirical way to distinguish between the strictly decomposed and the hybrid model based on their implications for number comparison. Another weakness is the qualitative nature of the models of two-digit number processing. Neither the holistic nor the decomposed or the hybrid model allow for quantitatively predicting the RT pattern in a magnitude comparison task. In all cases, the models can only make ordinal predictions about mean RT differences under different conditions (e.g., small vs. large distances between the to-be-compared numbers); however, it is not possible to quantify these differences. At this point, computational models would be informative as they offer the possibility to evaluate model predictions quantitatively. Therefore, the current study aimed at differentiating between the three models of two-digit number processing described above via a computational modelling approach validated by empirical data from number magnitude comparison. In particular, we intended to distinguish between the strictly decomposed and the hybrid processing model by means of their modelled data as these two models cannot be differentiated by empirical RT/error data. Nevertheless, evaluation of the fit of the data produced by either of these models and the empirical data would be informative about the plausibility and validity of both strictly decomposed as well as hybrid processing of two-digit numbers.

Before turning to modelling specifics, differences between the current study and the rationale behind other recent attempts to implement number magnitude representation into computational neural networks shall be discussed briefly.

Computational models of number magnitude representation

As already touched on before, computational models offer the possibility to evaluate the plausibility and validity of a proposed model not only qualitatively but also on a more fine grain quantitative level. Quantitative computational models allow for a statistical appraisal of the fit between modelled and empirical data – thereby, qualifying a more objective validation of models predictions. For the case of number processing, most previous computational models of number magnitude representation incorporated some kind of number line assumption (e.g., McCloskey & Lindemann, 1992; Viscuso, Anderson, & Spoehr 1989, Zorzi & Butterworth, 1999). Following this conceptualization each number is represented by a single node in an ordered sequence of input nodes. In several studies by different authors and model architectures it was observed that empirical data from different numerical tasks was associated reliably with the modelled data (number comparison: e.g., Dehaene & Changeux, 1993; Zorzi & Butterworth, 1999; number naming: e.g., Grossberg & Repin, 2003; Verguts et al., 2005; number priming: Verguts, Stevens, & Fias, 2005; Zorzi, Stoianov, & Umiltà, 2005). However, almost all of these models were focused on small numbers and numerosities up to 15 (Verguts et al., 2005; up to 5 in Dehaene & Changeux, 1993; 9 in Zorzi & Butterworth, 1999; but see the discussion of the Grossberg & Repin, 2003 model below). Moreover, considering the relevant literature revealed that most of the papers on computational models of number representation were concerned with questions regarding the actual coding characteristics upon the assumed and implemented number line representation. Recently, summation or numerosity coding (i.e., representing magnitude by the number of nodes activated, e.g., Zorzi & Butterworth, 1999) and place coding (i.e., representing a specific number by the activation of a node actually reflecting its position on the number line as well as the preceding and subsequent one, e.g., Verguts & Fias, 2004; Verguts et al., 2005) were evaluated. Neither of these coding schemes differentiates between the representations of single-digit and two-digit numbers: in the first case all nodes along the number line up to the i^{th} node are activated to code magnitude i and in the latter case only the i^{th} node is activated to do so. In both cases no differentiation between e.g., tens and units for two-digit numbers is assumed (but see Grossberg & Repin, 2003). Taken together, to date computational modelling was primarily employed to clarify coding and scaling aspects of numerical magnitude along the mental number line rather than distinguishing between different processing models (see Verguts et al., 2005 for a more detailed discussion of this point).

However, the model by Grossberg and Repin (2003) presents an exception to these considerations. In this model, representations of single- and multi-digit numbers are discerned. The authors assume multi-digit numbers to be represented by an interaction of the ventral “what” and the “dorsal” where stream of processing visual stimuli: In this way the one dimensional representation of single-digit numbers is augmented by a second dimension coding the base-10 property of the Arabic number system. However, Grossberg and Repin (2003) did not specify whether this two-dimensional coding of multi-digit numbers also manifests itself in separate representations of e.g., tens and units or rather converges on one holistic representation, upon which the comparison process is based. Nevertheless, inspection of Figures 22B as well as 23A (Grossberg & Repin, 2003, p. 1130) provides at least descriptive evidence suggesting compatibility like effects for both error and RT data produced by the model. From the distribution of relative errors committed by the model (see Figure 22B, Grossberg & Repin, 2003, p. 1130) one can derive that for number smaller as well as larger than the standard (i.e., 55) incompatible comparisons (e.g., 38_55 or 55_72) were associated with more errors than compatible comparisons with even smaller overall distance (e.g., 41_55 or 55_69, see also Dehaene et al., 1990 for a discussion about these discontinuities, i.e., reversed distance effects at decade boundaries). A similar effect is evident for the original simulated RT data (mainly for probes smaller than the standard) with the effect being most pronounced at the transition from the second to the third decade (see Figure 23A, Grossberg & Repin, 2003, p. 1130). Taken together, these observations implicate that the model by Grossberg & Repin (2003) may be based on the assumption of decomposed processing of tens and units but in the paper there is no explicit statement on this issue. Most importantly for the current study, Grossberg and Repin (2003) did not aim at differentiating between the strictly decomposed and the hybrid processing approach.

The present study

Nevertheless, in the present study we were not interested to further add to the ongoing debate about possible coding and scaling characteristics of the mental number line and their implementation in computational models. Instead, we aimed at employing computational modelling to evaluate the plausibility and validity of above introduced models of two-digit number processing (i.e., holistic, decomposed, and hybrid processing) for the case of two-digit number comparison. So far, there is ample evidence for decomposed processing of two-digit numbers (e.g., Nuerk et al., 2001; Moeller et al., 2009a). However, at the moment there is no empirical variable allowing a reliable differentiation between the strictly decomposed and the hybrid model. Therefore, particular interest was paid to the question whether the assumption of an additional holistic representation of overall magnitude as proposed by the hybrid model (cf. Nuerk & Willmes, 2005) is actually necessary to account for the empirical data. This question is specifically interesting as the hybrid model is the most complex one and following the principle of parsimony/theoretical economy it should only be chosen when the other models and in particular the strictly decomposed model (cf. Verguts & De Moor, 2005) do not account for the data at hand.

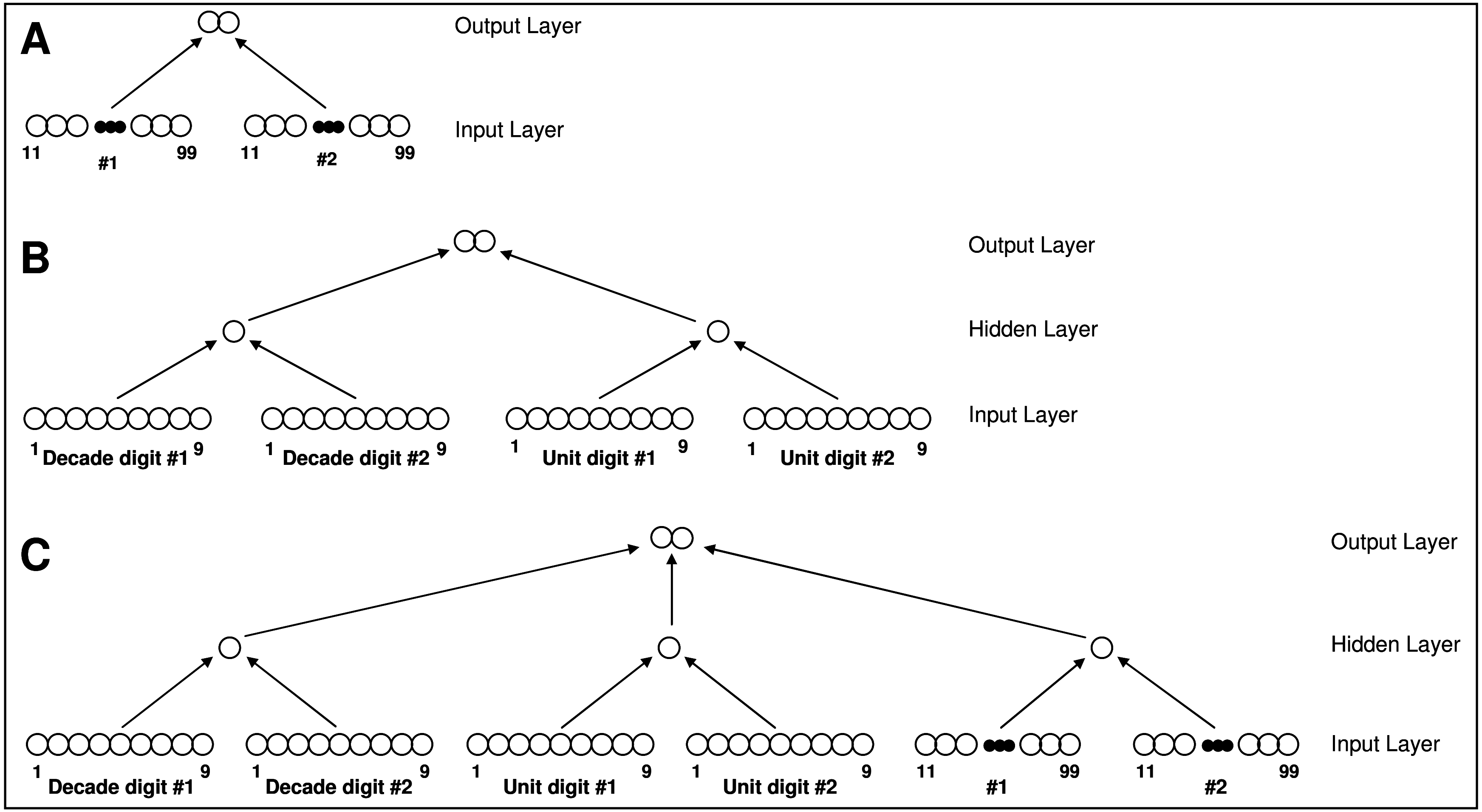
In summary, the current study set off to distinguish between three proposed models of number magnitude representation (i.e., holistic, decomposed, and hybrid model) in a computational modelling approach. In a first step, the validity of the computational models is evaluated by their ability to account for the distance effect, a standard effect observed in numerical cognition research. It is expected that all three models should be able to account for the distance effect. In a second step, it is aimed to evaluate the influence of decomposed processing of tens and units as indicated by the compatibility effect. Here, it is hypothesized that the holistic model should fail to account for the unit-decade compatibility effect while both the strictly decomposed and the hybrid model should be capable of accounting for the compatibility effect. Third, it shall be appraised whether incorporation of a predictor reflecting a measure of overall distance [i.e., (logarithmic) overall distance or the distance between the logarithms of the to-be-compared numbers] necessarily requires a holistic magnitude representation as previously argued by Nuerk and Willmes (2005, see also Nuerk et al., 2001; Knops, 2006). If so, such a predictor should only be included in the final regression model evaluating the data produced by the hybrid model, but not in the regression model for the strictly decomposed data as the latter model does not incorporate a representation of overall magnitude. Finally, following Occam's razor the hybrid model should only be chosen as the model predicting the empirical data in the best way when its performance is found to be reliably superior to the performance of the strictly decomposed model which is the more parsimonious one.

In the next section the architecture and the implementation of the three models will be described before actually reporting the models' performance in two-digit number comparison.

METHODS

Model description

According to the model assumptions on the mental representation of two-digit numbers put forward to account for the empirical data, three different neural network models were programmed: (i) a holistic model in which to-be-compared two-digit numbers (i.e., indicated as #1 and #2) are represented as integrated entities (see Figure 1, Panel A), (ii) a strictly decomposed model with two distinct representations of tens and units (see Figure 1, Panel B) and (iii) a hybrid model in which the characteristics of the former two models are combined so that separate representations of tens and units operate in parallel with a holistic representation of overall magnitude of the respective number (see Figure 1, Panel C). The neural networks reflecting each of these models were programmed using MatLab 7.4.0.



<u>Figure 1:</u> Schematic layout of the architecture of the three computational models used in the current study: Panel A depicts the holistic model (e.g., Dehaene et al., 1990), Panel B illustrates the strictly decomposed model (cf. Verguts & De Moor, 2005), and Panel C reflects the hybrid model of two-digit number representation (e.g., Nuerk & Willmes, 2005).

The detailed description of the three neural network models will first focus on depicting the characteristics common to all three models before covering the specific features of each model in turn.

Features common to all three models

The number zero was not implemented in any of the three models as the mental representation of zero seems to be special as compared to the other single digits: For instance, Brysbaert (1995) observed that reading zero took participants reliably longer than reading any other single digit number. Moreover, Nuerk, Iversen and Willmes (2004b) provided additional evidence based on nonmetric multidimensional scaling analyses that the number zero may not be represented on the same number line as the other numbers. However, these studies cannot clarify whether the specificity of zero originates from the fact that zero occurs less frequently in daily life or from difficulties with the conception of a number symbol representing nothing (e.g., Frobisher, 1999; Levenson, Tsamir, & Tirosh, 2007). Taken together, it is still debateable whether the mental number line starts at zero or at one.

Finally, because all computational models of number magnitude representation start with a representation of one (e.g., Verguts et al., 2005; Grossberg & Repin, 2003; Zorzi & Butterworth, 1999) – zero was also left out in our modelling study. Basically there would be two different possibilities of how to implement zero, either by a separate node preceding the node associated with one or by no specific node at all. Because the latter alternative would result in no specific neural activation when semantically processing the number zero, this conceptualization seems problematic because the missing representation of zero would also result in missing representations of single-digit numbers and decade numbers: At least in the fully decomposed model a representation of zero is inevitable to account for numbers such as 04, 08, etc as well as 20, 30, etc, respectively. Please note that in contrast to the prolonged processing times observed for the number zero processing multiples of 10 is faster than expected considering their magnitude. To guarantee a valid representation of all numbers in all three models the number range implemented in the current models goes from 11 to 99 with exclusion of all multipliers of ten. The empirical data about the compatibility effect were also obtained without the employment of decade numbers. Nevertheless, modelling the role of zero in multi-digit numbers poses a challenge for the future.

The representation of number magnitude as assumed in the present study was identical to Verguts et al. (2005) in their models of small number representation. Each node represents one number, with numbers being aligned in ascending order from left to right. For instance, when presenting the number 5, the fifth node from the left gets maximally activated, with all preceding and adjacent nodes being co-activated symmetrically to some degree, depending on their distance to the maximally activated node. Nevertheless, activation always reaches its maximum of for the number actually presented.

All three models were programmed as feedforward networks with the decomposed as well as the hybrid model comprising an input layer, one hidden layer and an output layer. The holistic model is set up without a hidden layer (see below for a discussion). For all three models the hyperbolic tangent [a commonly used transfer function in neural network modelling, e.g., Harrington, 1993; Hérvas-Martínez et al., 2009; Ngaopitakkul & Kunakorn, 2009] was chosen as the activation function of the hidden layer and the output layer.

The propagation function is given in equation (1) with τ representing a model constant determining how fast the maximum activation of a node is reached and $\overline{net_j}(t)$ reflects the average input over time for a respective node *j*:

$$a_j(t) = \overline{net_j}(t) = \tau\, net_j(t) + (1-\tau) net_j(t-1) \qquad (1)$$

The output function of the hidden layer was the identity function making the output equal to the activation. The output of the output layer was determined as follows: The two nodes of the output layer reflect two distinct decisions. When the activation of the left node exceeds a threshold level θ, this indicates that the first number was the larger one of the pair. Contrarily, when the threshold θ is first exceeded by the right node, this means that the second number was larger. The threshold was set to $\theta = 0.5$ in all of the current networks (cf. Verguts et al., 2005; for a similar approach). Because activations approach a certain level in an asymptotic manner, there is the possibility that none of both activations exceeds θ and the loop would run forever. To prevent overly long “responses”, a time limit was introduced after which the decision process was terminated. The actual choice of a time limit is dependent on the choice of the parameter τ. For smaller/larger values of τ the loop takes longer/faster to reach maximum activation. A τ-value of 0.01 and a time limit of $rt_{\max} = 50$ loops were found to maximize the range of the simulated reaction time distribution while at the same time minimizing the number of time limits actually reached.

After the network model solved the comparison task for one pair of numbers within the given time limit, activation of the output nodes is compared to the correct output. For instance, when comparing 43 to 78, activation of the left node may be 0.12 and 0.67 for the right node. However, in this case, the correct output t_j for the left node would be 0 and 1 for the right node. Nevertheless, it is not mandatory that the obtained activations equal the correct output (cf. Verguts et al., 2005). For a correct decision the actual activation just needs to be closer to the true value for that particular output node than to the true value for the other output node, or put differently: in case the difference between the actual activation and the correct output does not exceed a difference of d > 0.5 the correct output is returned by the model. For our example, the output for the left node would be 0, because $|0.12 - 0| < 0.5$ and the output of the right node would be 1, because $|0.67-1|<0.5$, respectively. Generally, the output is calculated as follows:

$$o_j = \begin{Bmatrix} t_j & if\ |t_j - a_j| < d \\ a_j & if\ |t_j - a_j| \geq d \\ a_j & if\ rt = rt_{\max} \end{Bmatrix} \qquad (2)$$

Where t_j denotes the correct output and a_j reflects the actual activation of node j.

Differences between the models

The decomposed model comprises an input layer with four separate input arrays each representing one single digit. Going from left to right, the first input array reflects the decade digit of the first number, followed by the input array of the decade digit of the second number. The third input array represents the unit digit of the first number followed by the input array fort he unit digit of the second number (see Figure 1). These arrays are connected to a hidden layer, whereupon the decade digits project to one node and the unit digits project to another node of the hidden layer. At these nodes decade digits and unit digits of the two to-be-compared numbers are compared separately. The nodes of the hidden layer are then connected to an output layer in which the left node reflects the "first number larger" decision, whereas the right node is associated with the "second number larger" decision. Thereby the three layers can be associated with representations of the single digits (input layer), a comparison process (hidden layer), and a response field (output layer).

The hybrid model expands the decomposed model by a holistic comparison process. The input of the holistic part of the model comprises two arrays reflecting a holistic representation of the two to-be-compared numbers. These arrays project on a node in the hidden layer which in turn projects on the response related nodes in the output layer.

In the holistic model, two input arrays, each reflecting a holistic representation of the numbers ranging from 11 to 99, are implemented. These input arrays are connected to two output nodes, whereof the left one represents the “first number larger“ decision and the right one the “second number larger” decision.

Training of connection weights

The initial connection weights were pseudo-random values ranging from 0 to 1 taken from a uniform distribution. Basically, the same learning algorithm was used for all three models. Nevertheless, as the fully decomposed and the hybrid model involved a hidden layer learning followed a back-propagation approach in these two models. For the holistic model which did not incorporate a hidden layer the learning algorithm used reflected delta rule guided learning, respectively. Please note that back-propagation is a generalization of the delta rule approach to networks involving at least one hidden layer (cf. Rumelhart, Hinton, & Williams, 1986). Based on this and the fact that the same mathematical algorithm was used for the training of connection weights in all three models we are confident that differences in the learning algorithm did not determine the current results. Finally, the learning constant η was arbitrarily set to 0.1.

The training phase comprised 50.000 for all models. In each training circle two randomly chosen numbers between 11 and 99 (excluding multiples of 10) were presented to the network. The frequency of occurrence of each number during the training cycles was determined by its frequency of occurrence in daily life as assessed by taking the entries referenced in Google for the respective numbers as an estimate of their every-day occurrence (see Appendix A; Verguts & Fias, 2008 for a similar approach).

The procedure of the training phase will be described for the case of the decomposed model: At first, two numbers between 11 and 99 were randomly chosen with the exception of tie numbers (e.g., 44) and multiples of 10 as empirical studies showed anomalous processing advantages for these numbers (see above for a discussion). Then these numbers were split into their decade and unit digits. On the basis of these single digits the respective input pattern was computed. Then, the net activation of the hidden layer was determined. The hyperbolic tangent of this net weight was then used as the actual activation. The output function of the hidden layer was the identity function. The net activation of the output nodes was again computed by formula (1) with the hyperbolic tangent as the activation function. This was repeated until (i) the activation of one of the output nodes exceeded the threshold or (ii) the time limit was reached[1]. The actual output was then calculated using formula (2). Finally, the back propagation learning algorithm was applied meaning that the connection weights were adjusted on the basis of the difference between the activation of the output and the hidden layer of the network model.

The network model representing the holistic model was trained quite similarly, however, without breaking up the numbers presented into tens and unit digits. Nevertheless, as in this model the hidden as well as the output layer would consist of two nodes connected by an identity function (which would not alter the connection weights) no hidden layer was realized in this model. Thus, the holistic computational model comprised an input and an output layer only. Thereby, just one matrix of weights had to be learned, following the delta rule approach.

MODEL PERFORMANCE

For the case of brevity the way the connection weights developed while training the models will not be illustrated in the main text of this article. Nevertheless, for the interested reader Appendix B provides a detailed description of the development of the connection weights for each of the three models.

[1] Please note that for each of the three models the mean of the simulated RTs was more than 4 standard deviations from the upper time limit indicating that in the vast majority of trials a the neural network models had come to a decision far before the time limit was reached.

Modelling two-digit number comparison – analysis and results

Simulated reaction times produced by either of the models were analyzed according to the constraints of the studies of Nuerk et al. (2001) for comparisons to a variable standard (i.e., singling out the larger of two numbers) and Moeller, Nuerk, and Willmes (2009b) for comparisons to a fixed memorized standard and will be reported in turn. As the simulated RTs only ranged between 0 and 50 corresponding RTs on the millisecond scale were estimated by a linear regression approach. The slope and the intercept of the regression equation were computed by predicting the actual empirical RTs by the simulated RTs for the comparison to a variable standard (cf. Nuerk et al., 2001) and a fixed standard (Moeller et al., 2009b). All analyses reported subsequently will be conducted over item RTs. In the case of the empirical data this means the mean RT per item over all participants. However, in the case of the model data this reflects RTs per item averaged over 10 simulation rounds (cf. Verguts et al., 2005 for a similar proceeding). Analyses of both the model data concerning comparisons with a variable external standard and comparisons to a fixed standard involved a categorical AN(C)OVA as well as a multiple stepwise regression analysis (see Nuerk et al., 2001 for a similar procedure)[2].

The univariate ANOVA for comparisons to a variable standard incorporated the fixed factors data origin (empirical vs. modelled), unit-decade compatibility (compatible vs. incompatible number pairs), decade distance (small: 1 - 3 vs. large: 4 - 8), and unit distance (small: 1 - 3 vs. large: 4 - 8). The corresponding regression analysis included the predictors (i) overall absolute distance, (ii) logarithm of the absolute distance, (iii) distance of the logarithmic magnitudes, (iv) absolute problem size, (v) logarithmic problem size, (vi) unit distance (e.g., ranging from –8 for 49_61 to +8 for 51_69), (vii) absolute unit distance, and (viii) compatibility (see Nuerk et al., 2001; Moeller et al., 2009a for similar proceedings).

[2] Please note that results were identical when using the activation difference between the two output nodes instead of simulated RT as the dependent variable suggesting the present approach of simulating RTs to be valid.

Finally, a regression analysis only incorporating the four reliable predictors of item RT as identified by Nuerk and colleagues (2001; i.e., logarithmic distance, problem size, unit distance, and absolute unit distance) was conducted to evaluate in how far the predictors for empirical item RT also hold for modelled item RT.

For the comparisons with a fixed standard an ANCOVA was conducted to account for the influence of problem size which cannot be matched between compatible and incompatible comparisons when controlling for overall distance at the same time. Apart from the covariate problem size the ANCOVA discerned the factors unit-decade compatibility (compatible vs. incompatible) and unit distance (small: 1_3 vs. large: 4_6). Furthermore, a stepwise multiple regression analysis was run over item RTs including the following predictors: (i) overall absolute distance, (ii) logarithm of the absolute distance, (iii) distance of the logarithmic magnitudes, (iv) problem size, (v) logarithmic problem size, (vi) unit distance ranging from -6 for e.g., 71_57 to +6 for e.g., 53_69, (vii) absolute unit distance and (vii) compatibility[3]. Again, in a final step a regression analysis only incorporating the four reliable predictors of empirical item RT as identified by Moeller and colleagues (2009b; i.e., logarithmic distance and unit distance) was conducted.

Before turning to the results of the network models, the original results for the empirical RT data of the two original studies (i.e., Nuerk et al., 2001 and Moeller et al., 2009b) will be summarized briefly.

[3] Please note that classifying items according to these constraints would have resulted in a larger overall distance for compatible than for incompatible number pairs, thereby confounding the distinction between effects of overall distance and unit-decade compatibility. Thus, to balance overall distance between compatible and incompatible comparisons selectively chosen probes were excluded from further analyses in each of the two item croups. It is important to note that overall distance cannot be matched by excluding randomly chosen probes (see Moeller et al., 2009b for a more detailed discussion of this point). Additionally, as it is not possible match compatible and incompatible comparisons for both overall distance and problem size (½ * (standard + probe)), the latter was considered as a covariate in all analyses.

Comparisons to a variable standard

In the study originally reporting the compatibility effect Nuerk et al. (2001) observed a compatibility effect with response latencies following incompatible number pairs being reliably longer than latencies for compatible pairs (759 ms vs. 728 ms). Additionally, the compatibility effect was reliably moderated by unit distance as it was more pronounced for large unit distances as compared to small unit distances (52 ms vs. 12 ms; see Figure 2, Panel A).

Holistic model

ANOVA

On the one hand predicting empirical RTs on the basis of the holistic model data yielded a R^2 of .42, which was reliable [r = .65; F(1, 238) = 171.71, p < .001]. However, closer inspection revealed that important empirical effects cannot be accounted for by the holistic model. In the item ANOVA for the data produced by the holistic model there was only a reliable main effect of decade distance [F(1, 232) = 211.04, p < .001]. This indicated that according to the model data number pairs with large decade distance were responded to significantly faster than number pairs with a small decade distance (710 ms vs. 778 ms). Yet, neither the compatibility effect nor its interaction with unit distance were statistically reliable [main effect: F(1, 232) < 1; interaction: F(1, 232) < 1, see Figure 2, Panel B]. Moreover, no other main effect or interaction was found to be significant [all F < 1.59, all p > .21]. Taken together, this indicated that a reliable correlation of modelled and empirical RT data of its own is not a sufficient index of model validity.

Regression

The final model of the stepwise regression analysis on the model data produced by the holistic model accounted for a considerable amount of variance [adj. R^2 = .84, R = .92, F(4,

236) = 417.38, p < .001] and incorporated the three predictors: problem size, difference between the logarithmic magnitudes of the two to-be-compared numbers, and logarithm of problem size (see Table 1). Evaluation of the beta weights revealed that response latencies increased as (i) distance between the logarithms of the numbers decreased, (ii) absolute problem size as well as (iii) logarithmic problem size decreased. Thus, comparable to the results of the ANOVA no indication of an influence of unit-decade compatibility was observed in the RT data modelled by the holistic model.

Please note the unexpected direction of the influence of the predictors absolute and logarithmic problem size in the final model. However, the reversed problem size effect does not necessarily disprove the holistic model as it is probably driven by the interrelation of problem size and the difference between the logarithms of the two numbers. As explained in greater detail in Appendix C this interrelation is not linear as assumed in the regression analysis, but curvilinear instead meaning that up to a problem size of about 65, problem size is positively correlated with distance, which in turn drives the reversed problem size effect.

<u>Table 1:</u> *Predictors included in the final regression model for the holistic data*

Predictor	***B***	***b***	***t***	**sign.**	**Change in R^2**	**Raw correlation**	**Partial correlation**
Constant	1226.38	-	10.16	< .001	-	-	-
Inclusion of absolute overall distance					.50		
Problem size	-1.53	- .39	2.08	< .05	.33	- .55	- .13
Diff Log distance	- 216.83	- .74	23.91	< .001	.01	- .55	- .84
Exclusion of absolute distance					- .001		
Log problem size	- 185.89	- .37	1.99	< .05	.003	- .61	- .13

Finally, the multiple regression including the four reliable predictors of item RT identified by Nuerk et al. (2001) produced a model predicting modelled RTs reliably [adj. R^2 = .81, R = .90, F(5, 235) = 254.29, p < .001] and without a substantial loss of descriptive adequacy. However, when looking at the beta weights of the individual predictors it was found that only the influence of logarithmic overall distance [constant = 1060.03; B = -126.03; b = - .72, t = 25.29, p < .001] and problem size [B = -2.25; b = - .58, t = 20.53, p < .001] was significant. Item RTs increased as logarithmic overall distance and problem size decreased. Contrarily, the influence of unit distance [B = 0.26; b = .02, t = 0.72, p = .47] as well as absolute unit distance [B = 1.24; b = .05, t = 1.67, p = .10] was not reliable. Again, these results indicated item RTs as produced by the holistic model not to be determined by unit-decade compatibility, but rather by measures of overall distance and problem size.

Decomposed model

ANOVA

Corroborating our expectations predicting the empirical RTs by the modelled RT data showed a reliable R^2 of .50 [r = .71; F(1, 238) = 241.13.99, p < .001] indicating a shared variance of 50%. More particularly, analyzing the RT data produced by the strictly decomposed model revealed a reliable main effect of compatibility [F(1, 232) = 23.68, p < .001] with response latencies following a compatible number pair being significantly shorter than latencies following a incompatible pair (735 ms vs. 753 ms, respectively). Moreover, this main effect was moderated by the interaction of compatibility and unit distance in the expected direction [F(1, 232) = 6.28, p < .05, see Figure 2, Panel C]: the compatibility effect was reliably more pronounced for number pairs with a large unit distance as compared to pairs with a small unit distance (28 ms vs. 9 ms, respectively). Thereby, these results clearly corroborate the notion of decomposed processing of tens and units. Finally, the main effect of decade distance indicated that number pairs with a large decade distance were responded to faster than number pairs with a small decade distance [699 ms vs. 789 ms, respectively; F(1, 232) = 573.17, p < .001]. No further main effect or interaction was statistically reliable [all F < 1.35, all p > .25].

Regression

The stepwise regression analysis on the modelled RT data of the strictly decomposed model resulted in a final model including the predictors difference between the logarithms of the to-be-compared numbers and unit distance [adj. R^2 = .91, R = .95, F(3, 237) = 1199.20, p < .001, see Table 2]. Closer inspection of the beta weights indicated that reaction times increased as the difference between the logarithms of the two numbers decreased and, most importantly, response latencies also increased as the unit distance decreased. As unit distance is negative for incompatible number pairs but positive for compatible pairs this clearly indexed an influence of unit-decade compatibility on item RTs estimated by the strictly decomposed model.

<u>Table 2:</u> *Predictors included in the final regression model for the decomposed data*

Predictor	***B***	***b***	***t***	**sign.**	**Change in R^2**	**Raw correlation**	**Partial correlation**
Constant	834.62	-	354.23	< .001	-	-	-
Difference of logarithms	-303.40	- .94	48.20	< .001	.89	- .94	- .95
Unit distance	- 1.84	- .13	6.90	< .001	.02	- .17	- .41

Again, a multiple regression incorporating the four significant predictors of item RTs as observed by Nuerk et al. (2001) accounted for a considerable amount of variance [adj. R^2 = .85, R = .92, F(5, 235) = 325.56, p < .001] without a substantial loss of predictive power. Moreover, inspecting the beta weights revealed that apart from logarithmic overall distance [constant = 938.76; B = -168.35; b = - .87, t = 34.18, p < .001] and problem size [B = 0.86; b = .20, t = 7.93, p < .001], unit distance [B = -2.02; b = - .15, t = 5.78, p < .001] reliably predicted item RT in the expected direction. Item RT increased as logarithmic overall distance decreased, problem size increased and, most importantly, as unit distance increased; thereby, indicating a significant influence of unit-decade compatibility on Item RT. On the other hand, absolute unit distance was not a reliable predictor of item RTs [B = 1.27; b = .04, t = 1.73, p =

.09]. Taken together, analyzing the data produced by the strictly decomposed model provided further evidence for the validity of the notion of decomposed processing.

Hybrid model

ANOVA

Generally, the results for the data produced by the hybrid model were quite similar to those observed for the data of the strictly decomposed model. Predicting the empirical RTs on the basis of the modelled RT data resulted in a R^2 of .40 [r = .63; $F(12, 238) = 159.29$, $p <$ .001]. Moreover, the ANOVA over items showed a significant main effect of compatibility indicating estimated RTs for compatible number pairs to be reliably faster than RTs for incompatible number pairs [734 ms vs. 754 ms, respectively; $F(1, 232) = 26.70$, $p < .001$, see Figure 2, Panel D]. Additionally, the expected interaction of compatibility and unit distance was present in the data [$F(1, 232) = 11.65$, $p < .01$]. So, the compatibility effect was larger for number pairs with a large unit distance than for number pairs with a small unit distance (34 ms vs. 7 ms, respectively). Again, of the remaining contrasts and interactions only the main effect of decade distance [$F(1, 232) = 322.74$, $p < .001$] was significant despite the marginally significant interaction of compatibility and decade distance [$F(1, 232) = 3.39$, $p = .07$; all other $F < 1$]. This indicated that number pairs with a large decade distance were followed by shorter RTs as compared to number pairs with a small decade distance (708 ms vs. 779 ms, respectively). Additionally, the compatibility effect tended to be stronger for number pairs with a large decade distance than for pairs with a small decade distance (26 ms vs. 13 ms, respectively). .

Regression

For the RT estimates produced by the hybrid model the stepwise multiple regression analysis yielded an adjusted R^2 of .76 [R = .87, $F(6, 234) = 149.79$, $p < .001$]. The final model incorporated the following five predictors: the difference of the logarithms of the two to-be-compared numbers, unit distance, logarithmic problem size, absolute overall distance as well as absolute problem size (see Table 3). Inspecting the beta weights showed that modelled RTs increased with (i) decreasing difference between the logarithms of the two numbers, (ii) decreasing logarithmic problems size (iii) increasing overall distance, and increasing absolute problem size. Finally, in line with the assumptions on unit-decade compatibility estimated RTs also increased as unit distance decreased. Please note that the unexpected positive interrelation of modelled RTs with absolute distance may be driven by intercorrelations of the predictors included in the final model. Looking at the raw correlation reveals an initial influence in the expected direction.

<u>Table 3:</u> Predictors included in the final regression model for the hybrid data

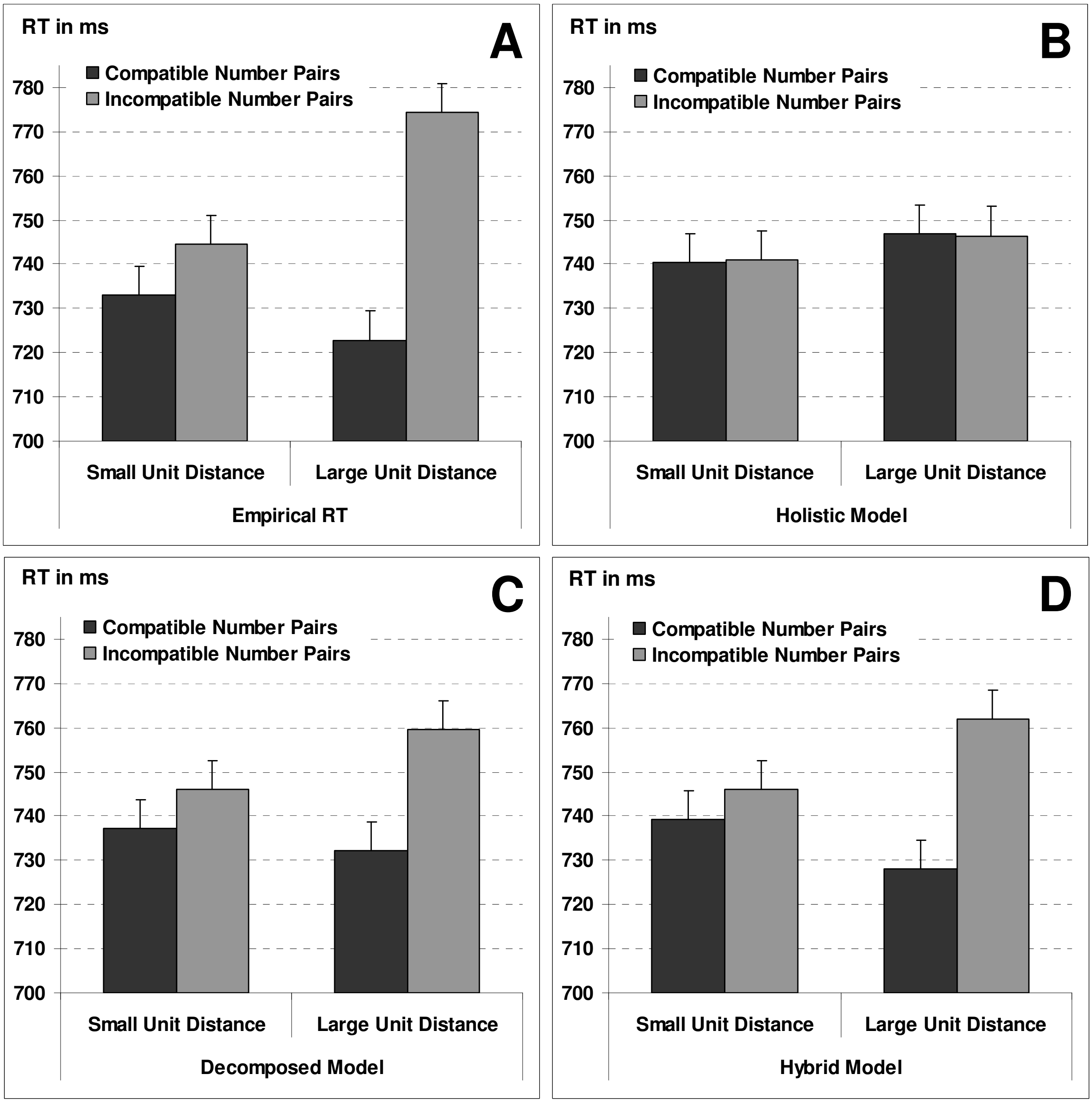
Predictor	***B***	***b***	***t***	**sign.**	**Change in R^2**	**Raw correlation**	**Partial correlation**
Constant	1761.05	-	11.07	< .001	-	-	-
Difference of logarithms	- 503.53	-1.75	9.31	< .001	.63	- .80	- .54
Unit distance	- 2.66	- .22	6.76	< .001	.05	- .25	- .40
Log. problem size	- 657.56	-1.33	5.46	< .001	.04	- .03	- .34
Abs. overall distance	2.46	.99	5.57	< .001	.03	- .77	.34
Abs. problem size	3.40	.90	3.79	< .001	.01	.05	.24

At last, the regression analysis including the four reliable predictors of item RT obtained by Nuerk et al. (2001) resulted in a significant model [adj. R^2 = .62, R = .79, F(4, 235) = 99.81, p < .001]. Nevertheless, compared to the stepwise regression analysis R2 was reduced considerably. Looking at the beta weights showed that item RT increased significantly when logarithmic overall distance [constant = 933.29; B = -129.60; b = - .75, t = 18.93, p < .001] as well as unit distance [B = -2.83; b = - .23, t = 5.82, p < .001] decreased, again indicating unit-decade compatibility to determine item RTs produced by the hybrid model. However, no reliable influence of problem size [B = 0.04; b = .01, t = 0.28, p = .78] and absolute unit distance [B = 0.22; b = .01, t = 0.21, p = .83] was observed. In summary, the data again indicated that decomposed representations of tens and units are necessary to account for the compatibility effect, even though they did not seem to preclude the existence of a concomitant representation of overall magnitude.

Contrasting the correlations with empirical data

Finally, to directly compare the individual correlation coefficients reflecting the interrelation between modelled and empirical data these correlation coefficients were transformed into Fisher's Z-values. Contrasting the Z-values showed that the data produced by the decomposed ($Z_{decomposed}$ = .89) exhibited a descriptively better fit to the empirical data than the holistic or the hybrid model which was however not statistically reliable ($Z_{holistic}$ = .78; Z_{hybrid} = .74; difference in z-scores: $Z_{decomposed}$ vs. $Z_{holistic}$ = 1.22; $Z_{decomposed}$ vs. Z_{yhbrid} = 1.59; critical difference at the .05 significance level = ±1.96). The difference between the model fit of the holistic and the hybrid model was also not statistically reliable (difference in Z-scores = 0.37).

In summary, the current computational results for comparisons to a variable standard indicated that (i) all three models were able to simulate the distance effect, (ii) the compatibility effect cannot be accounted for by a holistic representation of number magnitude, (iii) the inclusion of a predictor reflecting a measure of overall difference such as the difference of the logarithms of the to-be-compared numbers was not necessarily determined by the presence of a holistic representation either and (iv) the more complex hybrid model did not perform better than the more parsimonious strictly decomposed model.



<u>Figure 2:</u> Reaction times for compatible and incompatible number pairs separated for small and large unit distances. Panel A depicts the RT pattern as observed by Nuerk et al. (2001). Panel B shows the RT pattern as simulated by the holistic model, while Panels C and D reflect the results as found for the data produced by the strictly decomposed and the hybrid model, respectively. Error bars indicate 1 Standard Error of the Mean (SEM).

Model Comparisons to empirical date in a magnitude comparison task with a fixed standard

In their recent study Moeller et al. (2009b, i.e., Study 1 of the current thesis) presented evidence that the existence of decomposed processing of tens and units is not limited to an external representation of the to-be-compared numbers (cf. Zhang & Wang, 2005; Ganor-Stern, Pinhas, & Tzelgov, 2008). Instead, their results suggested that tens and units are represented separately also in the internal representation of two-digit numbers as the authors observed a reliable compatibility effect for comparisons of a given number to an internally memorized standard (RT for compatible comparisons: 492 ms vs. incompatible comparisons: 504 ms, see Figure 3, Panel A). Furthermore, this compatibility effect was found to be particularly driven by the compatibility effect for comparisons with a large unit distance (+ 16 ms) while it could not be obtained for comparisons with a small unit distance (+ 8 ms).

Holistic Model

ANCOVA

A regression analysis revealed that modelled RT data accounted for only 10% of the variance in the empirical RT data [$r = .32$; $F(2, 86) = 9.73$, $p < .01$]. Furthermore, analyzing the RT estimates produced by the holistic model showed that neither the main effect of compatibility nor its interaction with unit distance was statistically reliable [all $F < 1$]. Additionally, separate analyses for comparisons for large and small unit distances revealed that the compatibility effect was not significant in both analyses [both $F < 1$] indicating no modulation of the compatibility effect by unit distance (see Figure 3, Panel B). Nevertheless, the influence of the covariate problem size was significant in all analyses [two-way ANOVA: $F(1, 39) = 14.53$, $p < .001$; large unit distance: $F(1, 9) = 6.22$, $p < .05$; small unit distance: $F(1, 29) = 9.16$, $p < .01$]. Thus, as has been the case for comparisons to a variable standard (see above) the holistic model could not account for the unit-decade compatibility effect.

Regression

The final regression model for the data produced by the holistic model accounted for a significant amount of variance [adj. $R^2 = .31$, $R = .59$, $F(2, 37) = 9.67$, $p < .001$] and included the predictors absolute problem size and difference between the logarithmic magnitudes of the two to-be-compared numbers (see Table 4). Evaluation of the beta weights revealed that response latencies increased as distance between the logarithms of the numbers decreased and absolute problem decreased. As for the ANOVA results no indication of an influence of unit-decade compatibility was observed in the regression analysis on RT data modelled by the holistic model.

Table 4: Predictors included in the final regression model for the holistic data

Predictor	***B***	***b***	***t***	**sign.**	**Change in R^2**	**Raw correlation**	**Partial correlation**
Constant	566.18	-	40.46	< .001	-	-	-
Problem size	- .85	- .38	4.39	< .001	.25	- .55	- .13
Diff Log	- 90.63	- .36	2.28	< .05	.09	.02	- .35

Finally, the regression analysis incorporating only the two predictors found to reliably predict item RT by Moeller et al. (2009b) was predictive as well. However, the descriptive adequacy was considerably reduced as compared to the stepwise regression model [adj. R^2 = .16, R = .45, $F(2, 41) = 5.13$, $p < .01$]. Closer inspection of the beta weights revealed that only the influence of logarithmic overall distance on item RT was significant [constant = 543.12; B = -29.84; b = - .42, t = 2.99, p < .01], whereas no reliable influence of unit distance could be observed [B = 0.49; b = .17, t = 1.22, p = .23]. Item RT increased as the logarithmic distance between the to-be-compared numbers decreased. Taken together, the current analyses again indicate that the compatibility effect is hard to reconcile with the notion of a purely holistic representation of number magnitude.

Decomposed model

ANCOVA

For the interrelation of modelled and empirical RT it was observed that predicting empirical RT by the model data yielded a R^2 of .54 indicating a shared variance of 54% [r = .74; $F(2, 70) = 82.89$, $p < .001$]. Furthermore, the ANCOVA on the RT data produced by the strictly decomposed model revealed a reliable main effect of compatibility [$F(1, 39) = 6.03$, $p < .05$] meaning that estimated RTs for compatible comparisons were significantly shorter than estimated RTs for incompatible comparisons (496 ms vs. 506 ms, respectively). Moreover, both the main effect of unit distance and the interaction of compatibility and unit distance were not reliable [both $F(1, 39) < 1$]. Yet, when evaluating the compatibility effect separately for comparisons with either a large or a small unit distance it was found that the compatibility effect was present only for comparisons with a large unit distances [496 ms vs. 508 ms respectively; $F(1, 9) = 13.16$, $p < .01$] but missed significance for the comparisons with small unit distances [498 ms vs. 504 ms respectively; $F(1, 29) = 2.02$, $p = .17$, see Figure 3, Panel C]. Comparably to the analyses on the data produced by the holistic model the influence of the covariate problem size was reliable in all analyses [two-way ANOVA: $F(1, 39) = 15.27$, $p < .001$; large unit distance: $F(1, 9) = 7.71$, $p < .05$; small unit distance: $F(1, 29) = 10.79$, $p < .01$]. These results not only indicated that the strictly decomposed model accounted for a larger part of the variance of the empirical RTs in general, but also showed that the compatibility effect was simulated successfully.

Regression

The final model of the stepwise regression analysis on the modelled RT data of the strictly decomposed incorporated the predictors difference between the logarithms of the to-be-compared numbers, unit distance and logarithmic unit distance [adj. $R^2 = .80$, $R = .90$, $F(3, 40) = 59.49$, $p < .001$, see Table 5]. A closer look at the beta weights revealed that reaction times increased as the overall distance between the to-be-compared numbers decreased and problem size increased. More particularly, response latencies also increased as unit distance decreased. As unit distance is positive for compatible pairs but negative for incompatible number pairs the inclusion of this predictor reflects a reliable influence of unit-decade compatibility on response latencies estimated by the strictly decomposed model.

Table 5: Predictors included in the final regression model for the decomposed data

Predictor	***B***	***b***	***t***	**sign.**	**Change in R^2**	**Raw correlation**	**Partial correlation**
Constant	491.63	-	78.30	< .001	-	-	-
Overall Distance	-2.04	- .69	10.16	< .001	.48	- .69	- .85
Problem size	.74	.50	7.42	< .001	.25	.50	.76
Unit distance	- 1.14	- .29	4.33	< .001	.09	- .30	- .57

At last, the regression analysis including only the two predictors identified to significantly predict item RT in the study by Moeller et al. (2009b) resulted in a reliable model [adj. R^2 = .51, R = .73, $F(2, 41)$ = 23.23, p < .001]. However, descriptive adequacy of the model was considerably reduced as compared to the final model of the stepwise regression analysis. Nevertheless, both variables predicted item RTs reliably. Inspection of the beta weights revealed that item RTs increased as logarithmic overall distance [constant = 575.48; B = -63.81; b = - .66, t = 6.20, p < .001] as well as unit distance [B = -1.12; b = - .29, t = 2.68, p < .01] decreased. Again, the reliable negative influence of unit distance indicated item RTs to be driven by unit-decade compatibility. In summary, these analyses consistently indicated a good fit between the data produced by the strictly decomposed model and the empirical results, not only regarding descriptive adequacy, but also concerning the reliable simulation of the relevant effects (i.e., distance and compatibility effect).

Hybrid model

ANCOVA

Overall, predicting empirical RTs by the RT estimates produced by the hybrid model resulted in a R^2 of .43 [r = .65; $F(2, 70)$ = 52.24, p < .01]. However, unlike previously observed for the comparisons to a variable standard (see above) the strictly decomposed and the hybrid model did not perform similarly for comparisons to an internal standard. No reliable main effect of compatibility was present in the RT data produced by the hybrid model [$F(1, 39) = 1.24$, $p = .27$, see Figure 3, Panel D]. This indicated that modelled RTs for compatible comparisons were not reliably faster than RTs for incompatible comparisons (501 ms vs. 505 ms, respectively). Also, the interaction of compatibility and unit distance was not statistically reliable [$F(1, 39) < 1$]. In line with this, separate analyses for comparisons with large and small unit distances did not reveal the expected dissociation: neither for comparisons with a large unit distance [$F(1, 9) = 1.38$, $p = .27$] nor for comparisons with a small unit distance [$F(1, 29) < 1$] a reliable compatibility effect was found. This means that the hybrid model could not account for the compatibility effect as observed for comparisons to a fixed standard. Problem size as incorporated by a covariate did not exhibit a reliable influence in any ANCOVA [$F(1, 39) = 1.85$, $p = .18$; large unit distance: $F(1, 9) = 3.44$, $p = .10$; small unit distance: $F(1, 29) < 1$].

Regression

The only reliable predictor incorporated into the final regression model for the RT data modelled by the hybrid model was overall distance between the to-be-compared numbers [adj. $R^2 = .46$, $R = .69$, $F(1, 38) = 37.56$, $p < .001$]. Looking at the beta weight showed that reaction times increased as the distance between the two numbers decreased [constant = 533.22; B = -2.03; b = - .69, t = 6.13, p < .001].

Rerunning the regression analysis with incorporation of the two reliable predictors as identified by Moeller et al. (2009b) did result in a statistically reliable model without considerable loss of predictive power [adj. $R^2 = .42$, $R = .67$, $F(2, 41) = 16.39$, $p < .001$]. However, inspection of the beta weights showed that only the influence of the predictor logarithmic overall distance was significant [constant = 575.50; B = -62.72; b = - .66, t = 5.62, p < .001] while unit distance did not predict item RT reliably [B = - .43; b = - .11, t = 0.95, p = .35].

Taken together, these results revealed that the compatibility effect for comparisons to a fixed standard could not be simulated by the hybrid model.

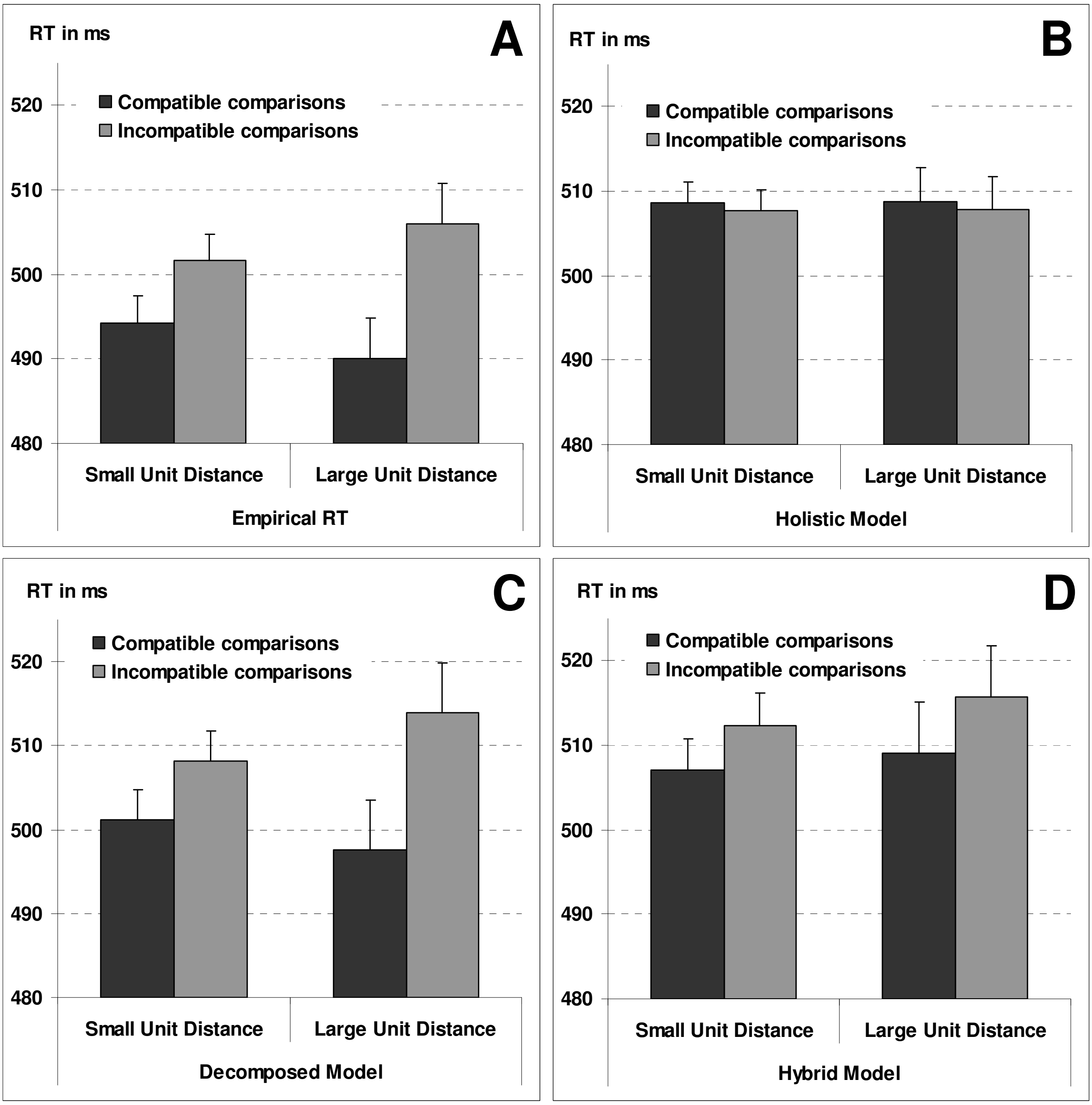


Figure 3: *Response latencies for compatible and incompatible comparisons to a fixed standard separated for small and large unit distances. Again, Panel A depicts the empirically observed RT pattern (Moeller et al., 2009b), whereas Panels B, C, and D indicate the RT pattern for the data produced by either the holistic, the strictly decomposed, or the hybrid model, respectively. Error bars reflect 1 SEM.*

Contrasting the individual models' correlations with empirical data

Finally, individual correlation coefficients reflecting the interrelation between modelled and empirical data were transformed into Fisher's Z-values to directly contrast them. As could already been inferred from the raw correlations the data produced by the decomposed and the hybrid model ($Z_{decomposed}$ = .95; Z_{hybrid} = .78) showed a significantly better fit to the empirical data than the holistic model ($Z_{holistic}$ = .001; difference in z-scores: holistic vs. decomposed = 5.40, holistic vs. hybrid = 4.50; critical difference at the .05 significance level = ±1.96). However, the difference between the model fit of the strictly decomposed and the hybrid model was not statistically reliable (difference in z-scores = 0.90).

In summary, the results pattern for the comparisons to a fixed standard only partly mirrored that observed for comparisons to a variable standard: (i) Again, all current computational models were able to simulate the distance effect. Additionally, (ii) the inclusion of the predictor overall difference in the final regression model was also not necessarily determined by the presence of a holistic representation. However, (iii) the compatibility effect could not be accounted for by the two models involving a holistic representation of number magnitude, i.e., the purely holistic and the hybrid model. Nevertheless, (iv) the strictly decomposed model did not perform better than the more complex hybrid model in terms of descriptive adequacy. Anyway, the data obtained from the strictly decomposed as well as the hybrid model achieved a satisfactory fit to the empirical data and performed reliably better than the holistic model.

DISCUSSION

The present study set off to provide a first comparison of the three currently proposed modes of two-digit number magnitude representations (i.e., holistic, strictly decomposed, or hybrid) using a computational modelling approach. It was expected that all three models should be able to account for the distance effect, whereas only the strictly decomposed and the hybrid model, but not the holistic model, should successfully simulate the unit-decade compatibility effect. Additionally, when incorporation of a predictor representing a measure of overall distance necessarily requires a holistic no such a predictor should be incorporated in the final regression model evaluating the data produced by the strictly decomposed model. Finally, due to reasons of model parsimony the hybrid model should only be chosen to be the most appropriate model when it outperforms the strictly decomposed model reliably.

Evaluation of the data produced by either of these models yielded results that were straightforward on all of these aspects. First, a reliable influence of the numerical distance between the to-be-compared numbers on RT was present in the data produced by each of the three computational models. Thereby, indicating that the distance effect was successfully accounted for by all three models for both, comparisons to a variable (as e.g., in Nuerk et al., 2001) as well as to a fixed standard (as e.g., in Moeller et al., 2009b; Study 1 of the present thesis). Second, as regards unit-decade compatibility our hypothesis was confirmed as the compatibility effect could not be observed for the data produced by the holistic model in general. Concerning the hybrid and the strictly decomposed model the results on the compatibility effect were less consistent. Although, the empirical data was accounted for comparably well by the strictly decomposed as well as the hybrid model in terms of descriptive adequacy, the compatibility effect could be simulated by both the strictly decomposed as well as the hybrid model only for the case of comparisons to a variable standard. Moreover, here, even the characteristic interaction of compatibility and unit distance was observed. The compatibility effect for modelled item RTs was consistently larger for number pairs with a large unit distance as compared to pairs with a small unit distance. Additionally, the categorical effect was specified by the results of the stepwise regression analyses for both the decomposed and the hybrid model. Unit distance was identified to be a reliable predictor of item RT for both models. Item RT increased as unit distance decreased. As unit distance is positive for compatible but negative for incompatible number pairs this again indicated the influence of unit-decade incompatibility on simulated item RTs.

Nevertheless, the results pattern was not so similar for the decomposed and the hybrid model when comparisons to a fixed standard were evaluated. Although both models did not differ reliably regarding descriptive adequacy, the compatibility effect was only found for the data produced by the strictly decomposed model. For the hybrid model data the compatibility effect was found in the expected direction, however, missed statistical significance. In line with this, the continuous predictor unit distance was a reliable predictor of item RT only in the regression analysis for the data produced by the strictly decomposed model. As for the comparisons to a fixed standard unit distance was limited to a maximum of ±6 and this maximal unit distance was present in only a small number of items, this may once again argue for the importance of large unit distances when looking for unit-based effects (see Moeller et al., 2009b) for a more detailed discussion (see also below for a discussion on possibly different origin of the compatibility effect within the strictly decomposed and the hybrid model). Taken together, this inconsistency in the replication of the compatibility effect between the decomposed and the hybrid model may be a first indicator of the superiority of the strictly decomposed model. Such an interpretation is corroborated by a general principle of model selection as well as another result of the current simulations.

On the one hand, taking into account that the hybrid model is the less parsimonious one and following Occam's razor the current computational data again suggest that multi-digit number magnitude is represented in a decomposed fashion retaining the place-value structure of the Arabic number system (see Myung & Pitt, 1997 for a discussion on model selection criteria). On the other hand, this interpretation can be further specified when considering the results of the regression analyses. Basically, the incorporation of a holistic representation of overall number magnitude in the hybrid model (Nuerk & Willmes, 2005) resulted from repeated findings that apart from predictors indicating decomposed processing of tens and units a measure of overall distance between the to-be-compared numbers (i.e., absolute distance, logarithmic distance, distance between the logarithms of the two numbers) was the most important predictor in a multiple regression analysis (e.g., Nuerk et al., 2001; 2002; 2004a; Moeller et al., 2009b; Wood et al., 2006). However, the current data indicated that a holistic representation (as in the purely holistic or the hybrid model) is not a mandatory prerequisite for an important influence of overall distance measures in the regression analyses. Instead, it was observed that even for the data produced by the strictly decomposed model a measure coding overall distance (i.e., distance between the logarithms of the to-be-compared numbers) was identified to be the most important predictor of item RT. Thus, this means that the inclusion of such a predictor is *not* a direct indicator for the existence of a holistic representation of number magnitude as it was also found for data produced by a model without any integrated representation of two-digit number magnitude.

In summary, the present results indicated that the distance effect could be accounted for by each of the three models, whereas the compatibility effect required decomposed representations of tens and units as suggested by Nuerk et al. (2001; see also Nuerk & Willmes, 2005 for a review). Additionally, the reliable influence of an overall distance measure in the regression analyses was not a necessary consequence of a holistic representation. Thus, the current data provide first computational evidence for the assumption of two-digit number magnitude representation to be organized in compliance with the place-value structure of the Arabic number system, thereby representing each power of ten (i.e., units, tens, etc.) separately. Under consideration of model parsimony the empirical data were accounted for best by a strictly decomposed model realizing separate representations of tens and units.

Different origins of the compatibility effect in the hybrid and the strictly decomposed model

When inspecting the development of the connection weights (see Appendix B) in the strictly decomposed as well as the hybrid model there are similarities but also important differences. On the one hand, the output of both models was significantly driven by the digits at the tens position. However, as indicated by a reliable unit-decade compatibility effect observed for the modelled RT data as well as the significant influence of unit distance in the stepwise regression analyses there seemed to be a considerable influence of the unit digits as well. Modelled item RT was reliably longer for incompatible than for compatible number pairs. Additionally, the compatibility effect was more pronounced for large as compared to small unit distances. Finally, item RT increased as unit distance decreased. As unit distance is negative for incompatible number pairs (e.g., 37_ 62 → decade distance: 6 - 3 = 3; unit distance: 2 - 7 = -5) the latter result again corroborates the interpretation of the compatibility effect to be determined by the magnitude of the unit distance. Taken together, these findings suggest that the results pattern observed in the empirical data (cf. Nuerk et al., 2001) can be simulated by both the strictly decomposed as well as the hybrid model of two-digit number representation.

Nevertheless, a closer look at the development of the connection weights also showed that the influence of the digits at the unit position was not identical in the strictly decomposed and the hybrid model. In the decomposed model the compatibility effect is determined by specific activation differences of the unit digits for compatible and incompatible number pairs and thus in line with the considerations by Nuerk and Willmes (2005; see also Knops, 2006). Yet, for the hybrid model, the direct influence of the unit digits on the output was less emphasized. Instead, the connection weights for the unit digits between the input layer and the hidden layer of the hybrid model were not altered considerably by the learning procedure and remained by far smaller than the weights for the tens digits and overall magnitude. As the activation of the output layer is a combination of the activation of the hidden layer for the comparisons of tens, units, and overall magnitude this indicated a very small and unspecific influence of the unit digits on the output in the hybrid model. For the compatibility effect this implies that its successful simulation by the hybrid model for the case of comparisons to a variable standard may not be primarily driven by specific activation differences of the unit digits between compatible and incompatible number pairs. Rather, the compatibility effect may be due to systematic differences in the difference between the activations for overall magnitude and that for the tens digit. As the separate representations of tens, units, and overall magnitude were assumed to be linear in the current computational models the latter difference nevertheless provides an indirect measure of the unit distance. Thereby, the compatibility effect observed for the data produced by the hybrid model is not based on the direct comparisons of the unit digits of the two to-be-compared numbers but by a derivate of the comparisons of overall magnitude and the tens. However, this indirect influence of the unit digits on overall performance differs from the influence originally assumed by Nuerk and Willmes (2005) for the hybrid model.

In summary, evaluation of the connection weights indicated the compatibility effect to originate from different sources in the strictly decomposed and the hybrid model. Interestingly, only the compatibility effect observed for the data by the decomposed model seemed to arise from the separate comparisons of tens and units of the to-be-compared two-digit numbers. On the other hand, the compatibility effect found for the hybrid model data seemed to be the result of differences between the representations of overall magnitude and the tens digits. Thus, only the comparison process realized by the decomposed model is in line with the original assumptions of the hybrid model as promoted by Nuerk and Willmes (2005; see also Knops, 2006). A probable reason for the obtained pattern of result may be linear dependency between overall magnitude, tens magnitude, and unit magnitude when assuming linear coding of all of these symbolic representations.

As already mentioned in the introduction the only other neural network model accounting for the whole range of two-digit numbers is the one by Grossberg and Repin (2003). Please note that Grossberg and Repin (2003) postulate the origin of the representation of multi-digit numbers within their model to be driven by verbal number words. This means, that their model was not explicitly organized along place-value constraints but assumed that apart from a *primary number line* coding single-digit numbers (but not being necessarily limited to these) so-called *number categories* associated with verbal markers (such as e.g., the suffix -ty) form the basis for the representation of multi-digit numbers. Thereby, the number categories should be more or less language specific as e.g., the Basque number word system inherits a combination of base-20 and base-10 properties. Nevertheless, although their model seemed to produce compatibility like effects (cf. Figure 23, p. 1130) the authors did not evaluate these systematically and neither did they interpret these effects in terms of place-value integration as proposed by the current study. Taken together, the model by Grossberg and Repin (2003) should also be capable of accounting for the compatibility effect as long as the number words used for training the model retain the base-10 place-value structure of the Arabic number system.

Finally, another aspect of input format apart from the number word dependency in the Grossberg and Repin (2003) model may be worth considering. As the separation of tens and units is only possible in symbolic numerical notations but not in non-symbolic notation (e.g., dot pattern) decomposed processing of tens and units may be limited to symbolic presentations. On the other hand, the representation of overall magnitude is not tied to symbolic notation that closely. Thereof, the question arises whether the representation of overall magnitude is mandatorily symbolic and linear as well – an issue discussed in the following section.

The nature of the representation of overall magnitude

Basically, the successful simulation of empirical effect by the present models assuming place coding of numerical magnitude argues for the employment of a symbolic representation of the to-be-compared numbers (cf. Verguts et al., 2005). However, in principle, the representation of two-digit number magnitude may indeed be a combination of symbolic and non-symbolic representations with non-symbolic coding being reflected by summation- rather than place coding (e.g., Roggeman, Veguts, & Fias, 2007; see also Verguts & Fias, 2008 for a discussion). In particular, for the hybrid model one could argue that only the decomposed part of the model follows symbolic place-value constraints whereas the holistic part may rather represent some kind of non-symbolic approximate overall magnitude. Regarding model architecture such a conceptualization could be implemented by employing some kind of summation recoding for the overall magnitude representation. In the model of number representation by Verguts et al. (2005) such summation coding was assumed only for the case of non-symbolic input. However, the Zorzi and Butterworth (1999) model used summation coding for any kind of numerical input. Thus, there is the possibility that the representation of overall magnitude as proposed for the hybrid model of two-digit number representation may not be symbolic in nature but rather recodes the symbolic input into some kind of numerosity representation more similar to that assumed for non-symbolic quantities. Thereby, the hybrid model would be integrating numerical information in the most comprehensive way as both notation specific power as well as notation unspecific information can be considered. In this context, for instance, separate representations of digital magnitude complying with the power levels of the place-value structure of the Arabic number system would reflect notation specific representations. On the other hand, the numerosity representation reflecting overall magnitude of an involved number may be rather notation invariant. Taken together, under these considerations the hybrid model of number representation may open up a new way within the actual debate on whether number magnitude representation is notation specific (e.g., Cohen-Kadosh, Cohen-Kadosh, Kaas, Henik, & Goebel, 2007) or rather independent from input notation and thus amodal (e.g., Piazza, Pinel, Le Bihan, & Dehaene, 2007). This controversy was addressed in a recent target article by Cohen Kadosh and Walsh (2009). Therein, the authors argued for non-abstract notation specific representations of number magnitude in humans. However, as becomes evident from the commentaries by Cantlon, Cordes, Libertus, and Brannon (2009), Dehaene (2009), Ganor-Stern (2009), Grabner (2009) and others there is also considerable evidence suggesting an abstract notation invariant representation of number magnitude. Interestingly, Kucian and Kaufmann (2009) addressed this issue by claiming that this diverging pattern of empirical results may indicate that there are indeed notation dependent representations which nevertheless overlap to a certain degree, thereby mimicking an abstract notation independent representation of number magnitude. So, the question may not be either notation specific or notation unspecific but to what extend are notation specific and unspecific representations involved in human number processing capabilities? The latter notion was briefly mentioned by Cohen Kadosh and Walsh (2009, p. 322) but not elaborated on further. However, the hybrid model of number magnitude representation (Nuerk & Willmes, 2005) may offer a first theoretical framework for guiding future research on this question.

LIMITATIONS / PERSPECTIVES

When evaluating the computational results of the current study a number of presuppositions and limitations that have been taken when setting up the computational models should be kept in mind. Most importantly, programming of the current models was largely determined by the psychological implications of the theoretical models. This means that the separate representations of tens and units in the strictly decomposed model as well as the separate representations of tens, units, and overall magnitude have been kept apart by the present model architecture until their respective activations are evaluated at the output stage.

Thereby, possible interactions and interferences between the processing of the separate representations were not implemented into the current versions of the decomposed as well as the hybrid model. In turn, this rather minimalistic model architecture did not cover the entire complexity of the hybrid model, in particular, for which specific inhibitory and excitatory interrelations of the single representations are assumed (cf. Nuerk & Willmes, 2005). In this sense, the decomposed comparisons of tens and units as assumed in the current models do also not take into account possible comparison processes between the tens digit of the one number and the unit digit of the other or even between the two digits constituting one number which have been observed to occur when comparing two two-digit numbers (Wood, Nuerk, & Mahr, 2005). Taken together, it has to be noted that the results of the present study have been achieved using rather prototypical and simplified architectures for the three models of two-digit number representations to-be-compared. Nevertheless, these results suggested that computational modelling may be a fruitful methodological approach to evaluate the validity of theoretical models on the structure of two-digit number magnitude representation. Therefore, it is up to future studies to investigate whether the current findings also hold when more specific aspects of the individual models are implemented into the architecture of the computational models.

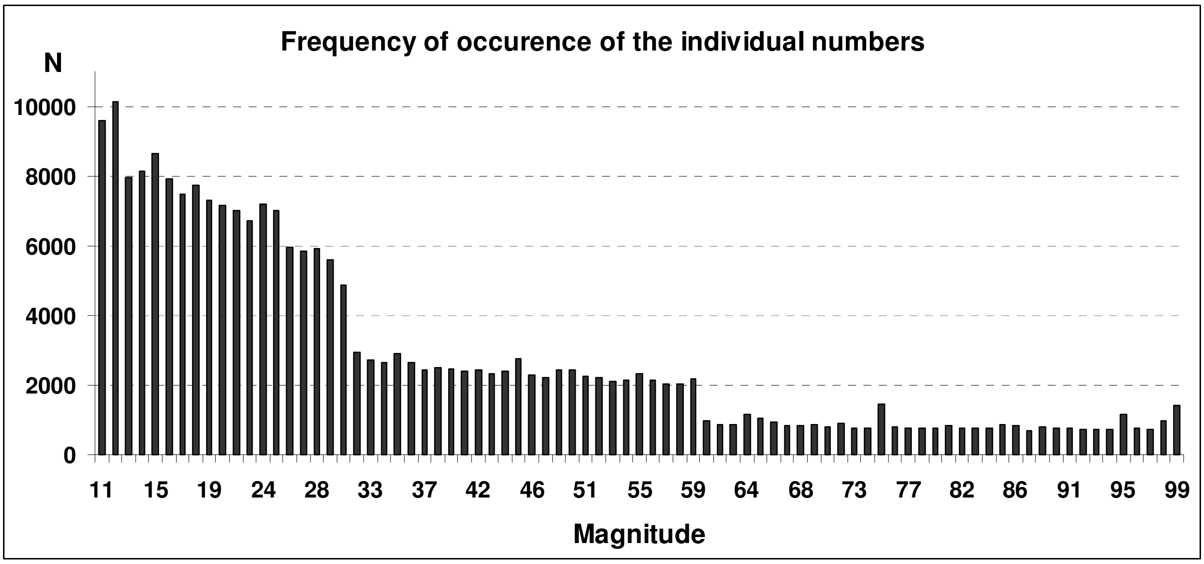
Another important point does not directly concern model architecture but more general aspects of cognitive processing. Nuerk and co-workers have argued repeatedly that attentional processes may play a role in two-digit number comparison (e.g., Nuerk & Willmes, 2005; Moeller et al., 2009b). In particular, the size of the compatibility effect was found to be influenced by the number of within-decade trials (e.g., 43_48) in the stimulus set. As in within-decade trials the unit digit is decisive increasing number of within-decade trials means that attentional focusing on the decade digits of the two to-be-compared numbers becomes less and less beneficial. This results in a more balanced relevance of tens and units for the decision and leads to a more pronounced compatibility effect. Additionally, in their current versions the strictly decomposed as well as the hybrid model assumes that processing of tens and units (and overall magnitude) works in perfect parallelism not allowing for any temporal gradient possibly delaying the processing of e.g., the units. At the moment the present computational models are not capable to account for such an effect of attentional modulation or any temporal specifications in the processing of the single representations. Again, future studies are required to investigate whether such processes indeed contribute to the processing of two-digit magnitude.

Taken together, it has to be noted that there are a number of questions still to be answered regarding both the implementation of processing specificities of the individual models as well as more general processing characteristics. Nevertheless, the current study indicated that computational modelling is a valuable tool for distinguishing between different models of two-digit number magnitude representation as it successfully discriminated between three theoretical models taking into account their capability to simulate standard effects observed in two-digit number comparison.

CONCLUSIONS

The aim of the present study was to provide first computational evidence for evaluating the validity of three current models of number magnitude representation: the holistic (e.g., Dehaene et al., 1990), the strictly decomposed (cf. Verguts & de Moor, 2005), and the hybrid model (e.g., Nuerk & Willmes, 2005). In particular, we were interested in distinguishing between the strictly decomposed and the hybrid model as there is currently no empirical way to discern the two models. The computational results were informative: As expected, the holistic model could not account for the compatibility effect whereas the decomposed and the hybrid model did. Moreover, a measure of overall distance between the to-be-compared numbers was the most important predictor of item RT even though the strictly decomposed model did not include an explicit representation of the overall magnitude of these two numbers. Synced with the fact that the decomposed model was the more parsimonious one the analyses indicated that the empirical data by Nuerk and colleagues (2001) and Moeller and co-workers (2009b) was accounted for best by the strictly decomposed model. In summary, the present study presented further evidence suggesting that two-digit number magnitude may not be represented holistically as has been claimed only recently (Ganor-Stern et al., 2008; Zhou, Chen, Chen, & Dong, 2008) but rather decomposed complying with the power levels of the place-value structure of the Arabic number system (e.g., tens and units).

APPENDIX A



<u>Figure A:</u> *Frequency of occurrence of individual numbers from 11 to 99 (excluding multiples of 10) as obtained from a Google survey and used to train the neural network models*

APPERNDIX B

Development of connection weights

Holistic model

For the holistic model training resulted in a diametrically opposed development of connection weights of the two output nodes for the two numbers: connection weights to one output node were comparatively high for small numbers and decreased with increasing magnitude of the number while this pattern was reversed for the connection weights to the other output node. Here, connection weights increased as the magnitude of the to-be-represented number increased. As can be seen from Figure 1 increase / decrease of connection weights with increasing magnitude did not follow a linear function, but rather seems to be of compressed (logarithmic) nature (see below for a discussion).

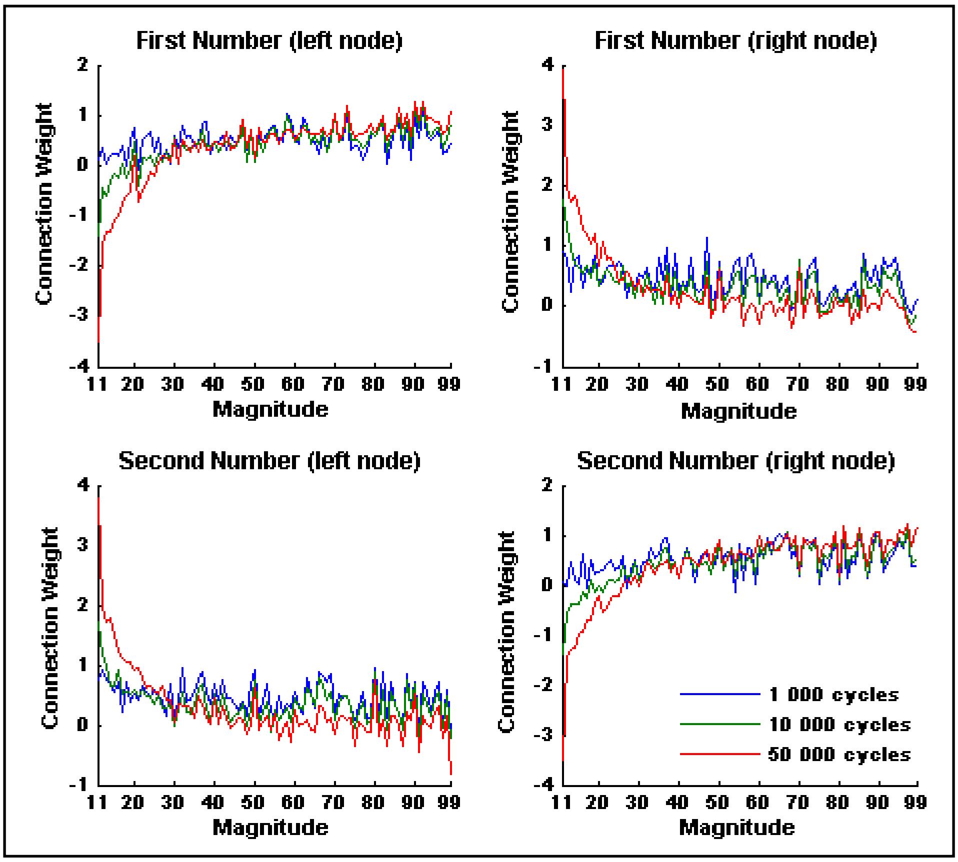
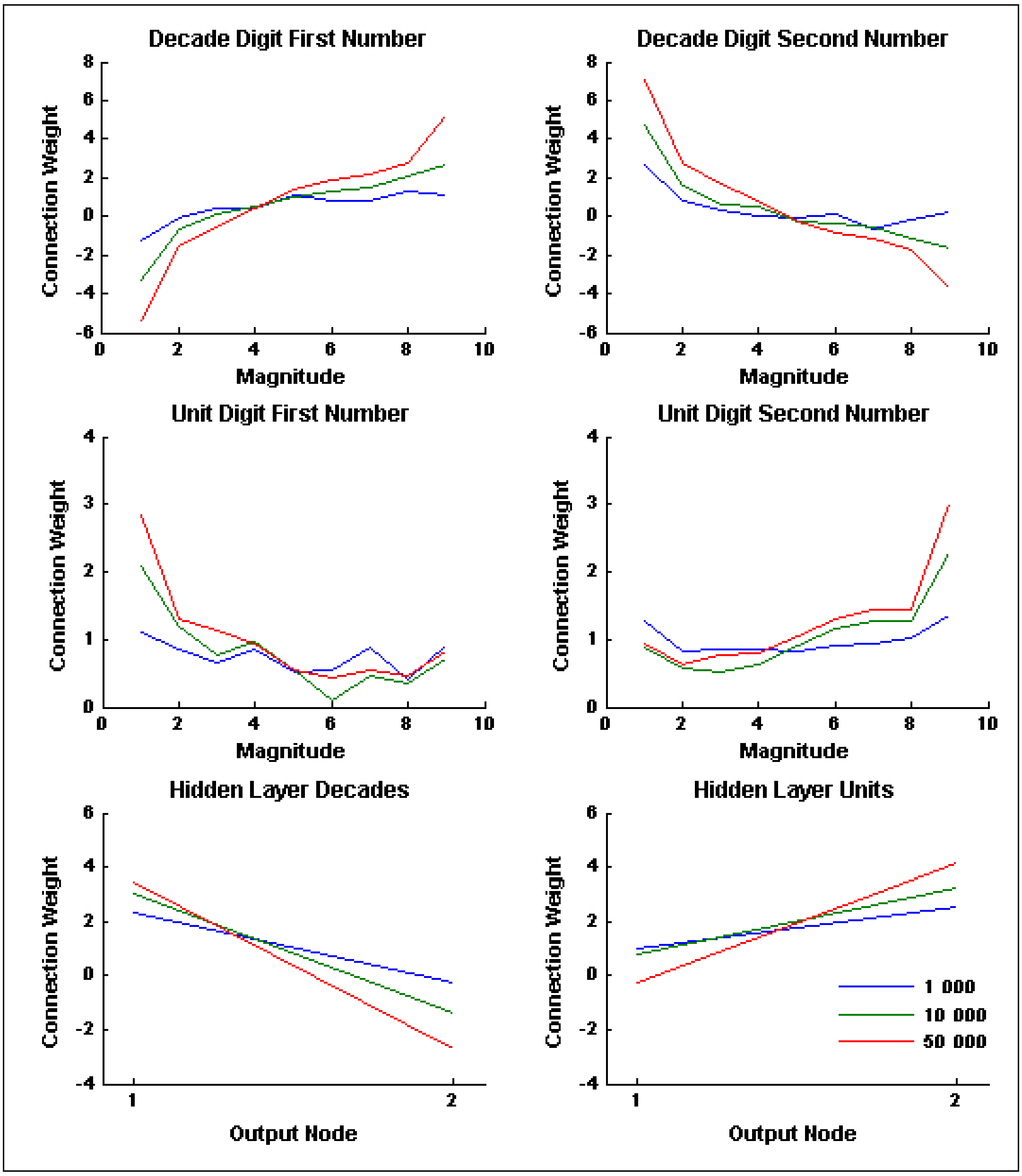


Figure B: *Connection weights from the input to the output node for each of the two to-be-compared numbers and each output node as observed for the holistic model (after 1.000, 10.000 and 50.000 training cycles).*

Decomposed model

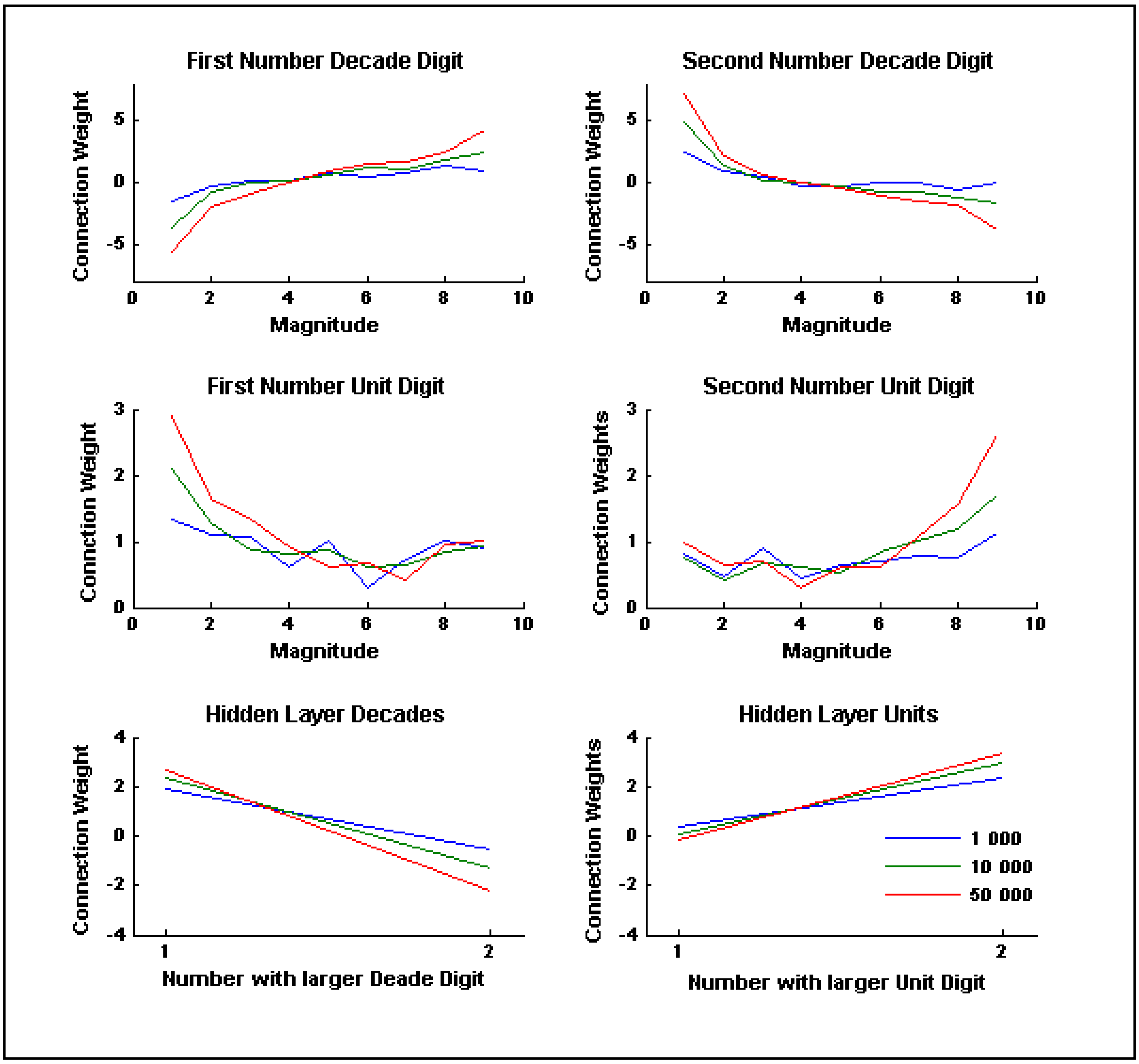
The connection weights of the two numbers (from the input to the hidden layer) as well as those for tens and units (from the hidden to the output layer) developed in diametrically opposed ways as described for the holistic model. Moreover, for the connection from the input to the hidden layer the connection weight pattern for decade and unit digits did not have a linear slope. Instead, as already observed in the model by Verguts et al. (2005) the pattern is described best by a compressed function as generally, differences between the connection weights of two successive numbers are larger for relatively smaller numbers (i.e., 2_3 vs. 7_8). According to Verguts and colleagues (2005) this pattern is determined by the frequency distribution of the single digits (i.e., 1 - 9) used for training the model, rather than allowing any kind of inferences on the scaling properties (i.e., logarithmic vs. linear) of the underlying magnitude representation. Importantly, in the present simulations the connection weights for the decade digits were stronger than the weights for the unit digits (with the difference between the minimum and maximum weight being about three times larger for the connection weights of the decade digits, see Figure C); thereby suggesting a more prominent influence of the decade digits in two-digit number magnitude comparison. This mirrors results repeatedly observed in regression analyses on empirical RT data. Here, the beta weights of for the predictor decade distance (when considered in the final model) were found to be two to four times higher than the beta weights for the unit distance (e.g., Knops, 2006; see Nuerk & Willmes, 2005 for a discussion).



<u>*Figure C:*</u> *Development of connection weights as observed for the strictly decomposed model (after 1.000, 10.000 and 50.000 training cycles). In the first row weights for the connection between input and hidden layer for the decade digits of the two to-be-compared numbers are given. The second row depicts the analogue for the unit digits. In the third row the connection weights between the hidden layer and the output layer are shown separated for tens and units.*

Hybrid model

The connection weights in the decomposed part of the hybrid model developed in a way similar to that in the strictly decomposed model. Again, the connection weights for the decade digits were higher than those for the unit digits. For the holistic part of the model the learning pattern mirrored the one found for the purely holistic model with the diametrically opposed development of connection weights for the two to-be-compared numbers. Importantly, for the connection between the hidden layer and the output layer the connection weights for the holistic part of the model developed similarly to the connection weights of the unit digits of the decomposed part of the model and diametrically opposed to the weights for



the decade digits.

Figure D: *Development of connection weights as observed for the decomposed part of the hybrid model (after 1.000, 10.000 and 50.000 training cycles). Again, the first row depicts connection between input and hidden layer for the decade digits while the weights for the units are given in the second row. The third row reflects the connection weights between hidden and output layer separated for tens and units.*

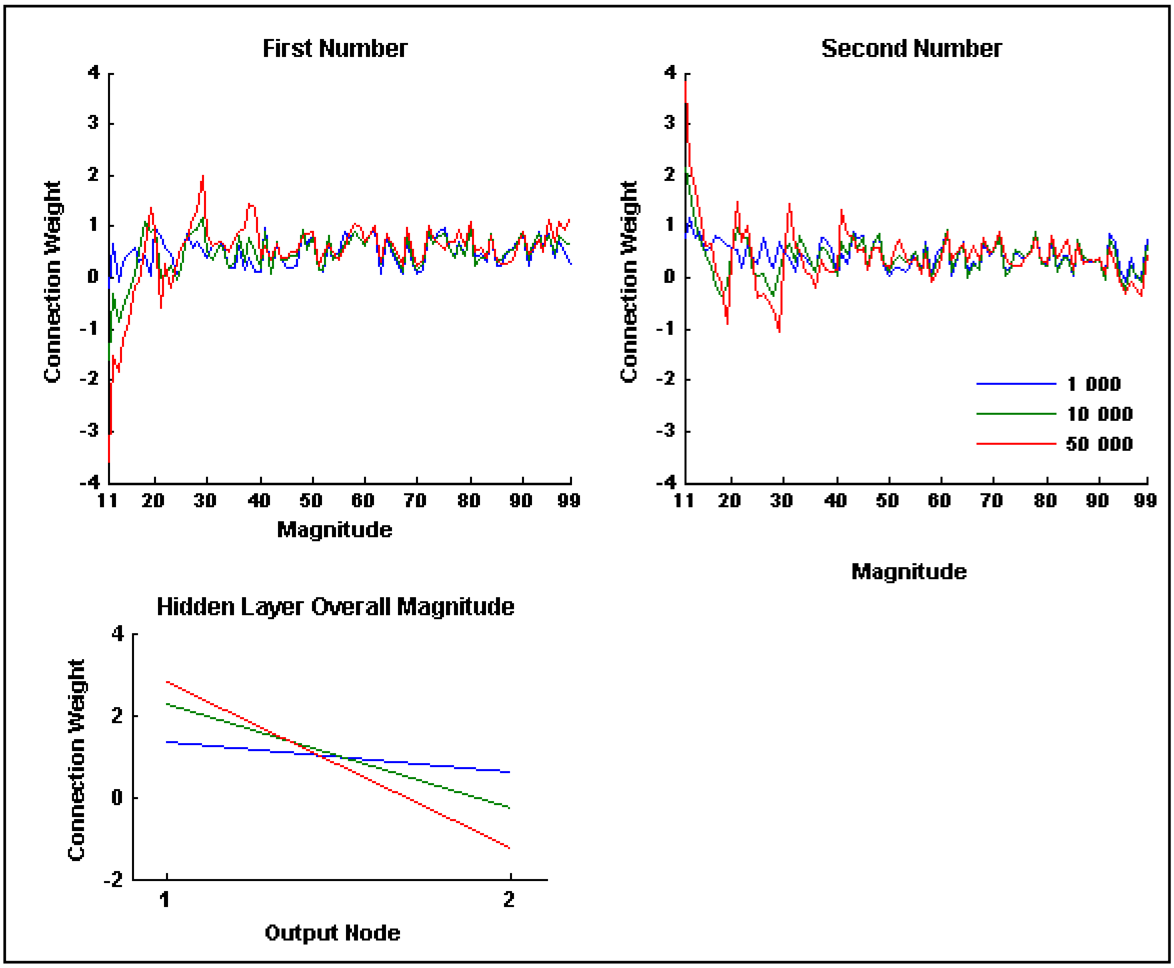


Figure E: *Development of connection weights for the holistic part of the hybrid model (after 1.000, 10.000 and 50.000 training cycles). The first row reflects the connection weights between the input and the hidden layer for each of the two numbers whereas in the second row the connection weights between the input and the output layer are depicted.*

Appendix C

On the interrelation of problem size and distance in the stimulus set used

As can be seen from Figure F the interrelation between the two reliable predictors of item RT produced by the holistic model (i.e., difference between the logarithms of the two numbers and problem size) is curvilinear rather than linear. However, the multiple linear regression analysis is not able to account for this. Closer inspection of the figure shows that the curvilinear relation means that up to a problem size of about 65 the difference between the logarithms is positively correlated with problem size. Only for items with a problem size larger than this the correlation becomes negative.

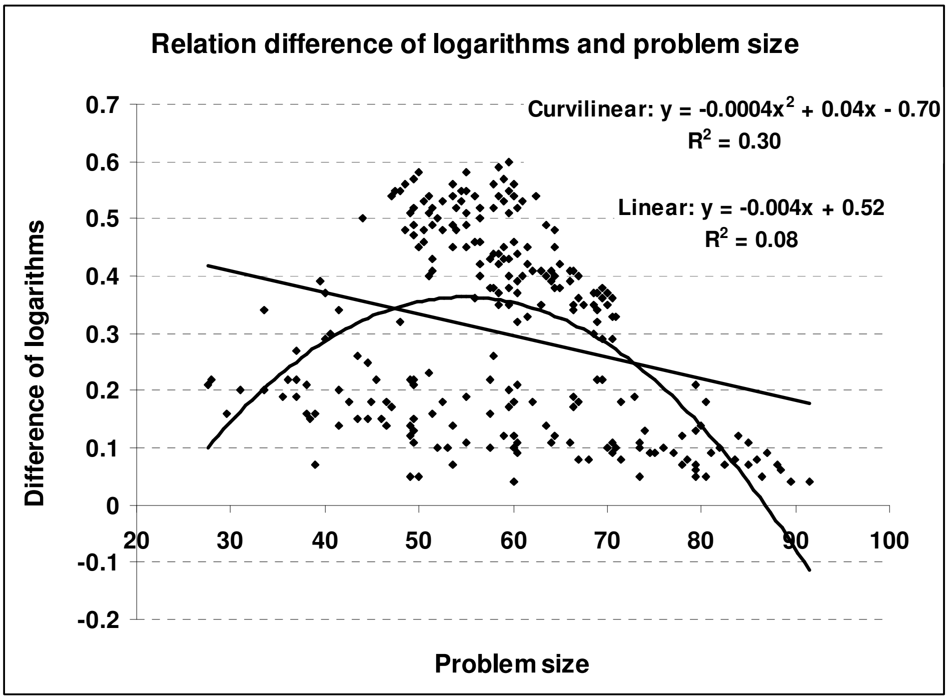
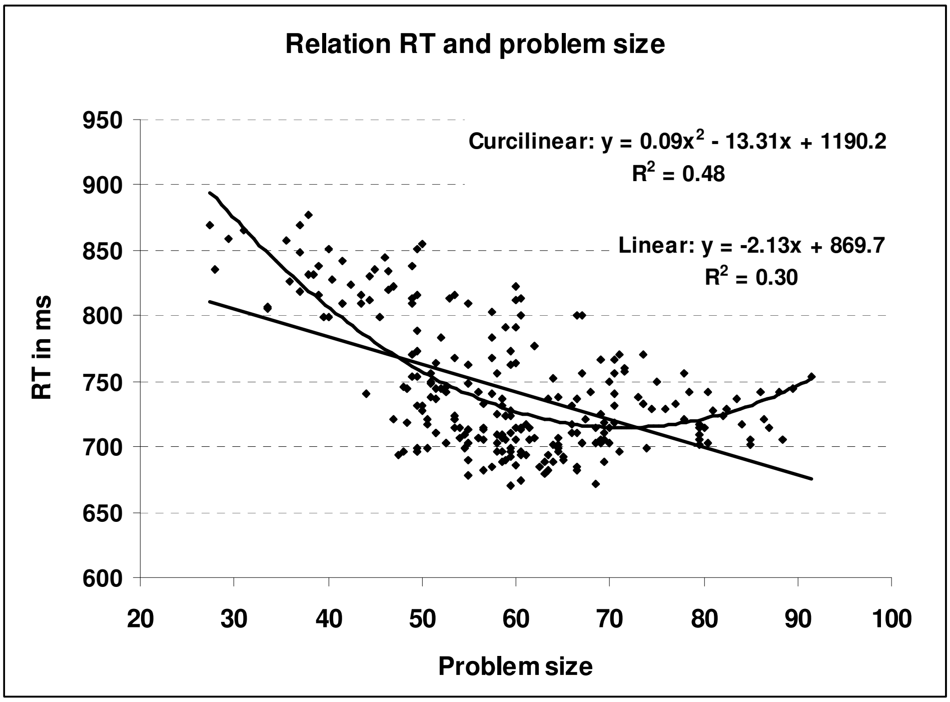


Figure F: *Simple linear and curvilinear fitting of the interrelation of the two reliable predictors of item RT difference between the logarithms and problem size*

At the same time, inspection of Figure G reveals that item RT decreases with increasing problem size up to exactly the same point of a problem size of about 65; thereby, indicating a reversed problem size effect. Only for items with a problem size above 70 a regular problem size effect of item RT increasing together with problem size can be observed.

Interestingly, one might infer from this that the distance effect dominates the problem size effect. However, at this point this conclusion has to remain speculative. Taken together, these additional analyses suggest that the unexpected negative influence of problem size on item RT may be determined by aspects of the stimulus set, rather than representing a shortcoming of the programming of the holistic model or the holistic magnitude representation in general.



<u>Figure G:</u> Simple linear and curvilinear fitting of the interrelation between problem size and item RT.

Section 6

Summary

SUMMARY

Before evaluating the implications of the preceding results in the General Discussion a short summary of each of the seven studies shall be given in turn to recapitulate the most important results. Please note that the summary will be focused on illustrating the most important results of the individual studies while the interpretation of the results will be kept at a minimum here and will be given in the Synopsis and Evaluation / Discussion sections of the General Discussion.

The current thesis aimed at investigating the influence of the place-value structure of the Arabic number system on numerical cognition in three different respects: (i) In a first set of studies questions regarding the cognitive representation / instantiation of the place value structure were addressed (cf. Section 2). (ii) Attention was paid to the importance of children's early place-value understanding and its developmental trajectories on the development of numerical / arithmetical competencies such as addition (cf. Section 3). (iii) The neuropsychological underpinnings of place-value processing were examined in a functional MRI study as well as in a patient study evaluating its impairments caused by hemi-spatial neglect (cf. Section 4). Apart from these empirical data this thesis provided a first evaluation of the representational structure of two-digit numbers (i.e., holistic, strictly decomposed, or hybrid) using a computational modelling approach (cf. Section 5). The summary will follow this structuring as the results of the respective sections will be illustrated in turn

Section 2: On task general influences of the place-value structure of the Arabic number system in human number representation

Study 1 of this thesis addressed the nature of the internal representation of two-digit number magnitude. Recent research (e.g., Zhang & Wang, 2005; Ganor-Stern, Pinhas, & Tzelgov, 2008) proposed that the decomposed representation of two-digit number magnitude may be determined by the presentational format employed in the respective studies. These authors argue that evidence suggesting separate processing of decade and unit digit magnitudes comes from studies in which both to-be-compared numbers were presented externally (e.g., 37_53). Additionally, they presented empirical evidence indicating holistic rather than decomposed processing of tens and units when only one number is presented and had to be compared to an internally memorized standard (Zhang & Wang, 2005). In the first study, when participants had to compare a given probe to an internal standard, empirical evidence for decomposed processing of two-digit number magnitude was observed as indicated by a standard unit-decade compatibility effect and its characteristic interaction with unit distance (i.e., a more pronounced compatibility effect for large unit distances). These results are hard to reconcile with the notion of a holistic internal representation of two-digit number magnitude in which the place-value structure of the Arabic number system is not retained.

In ***Study 2*** the influence of place-value information on the solving of basic arithmetic addition problems was investigated. It was hypothesized that place-value information is especially important in addition problems in which the execution of a carry operation is needed. In such trials the decade digit of the unit digit sum has to be carried to the result's decade digit to calculate the correct result, thereby, requiring processes of place-value, i.e., unit-decade integration. More particularly, it was expected that the basic processes of (i) calculating the unit sum to determine whether a carry is needed or not, and (ii) the updating of the decade digit of the result should be dissociable by evaluating participants' eye fixation behaviour. Based on the eye-mind and immediacy assumption (cf. Rayner & Pollatsek, 1989), a more pronounced increase in the number of fixations on the unit digits of the summands should be found if the difficulty of carry addition problems arises from computing the two-digit sum of these unit digits. Otherwise, if updating the decade digit of the result causes the difficulty of carry addition problems, there should be a steeper increase in the number of fixations on the result's decade digit. In line with these expectations, evaluation of the eye fixation data showed that the increase in number of fixations due to the carry operation was indeed more pronounced on the unit than on the decade digits of the summands. The reversed pattern was found for decade and unit digit of the result. In summary, the results corroborate the notion of carry addition problems being associated with specific processes of place-value integration.

Section 3: On the importance of place-value integration for arithmetic development

Study 3 set off to investigate the differential influence of children's early place-value knowledge on the development of their spatial representation of number magnitude. In a recent study, Moeller, Pixner, Kaufmann, and Nuerk (2009) found evidence for children's spatial number magnitude representation to be influenced by the place-value structure of the Arabic number system. Their estimates when localizing the position of a given number on a hypothetic number line are accounted for best by a two-linear model with a breakpoint at 10 representing two different representations for single- and two-digit numbers. The validity of this finding was evaluated in a cross-cultural study with German- and Italian-speaking children using a number word system with and without inversion, respectively. The results clearly indicated that the importance of the place-value structure of the Arabic number system for the development of a spatial representation of number magnitude was not determined by properties of the number word system. Similar to the German-speaking children, Italian-speaking children's estimation performance was also conceptualized best by a two-linear model reflecting two different representations for single- and two-digit numbers. However, estimations of German-speaking children were less accurate and in particular so on items for which an inversion error resulted in a large difference between the actual and the wrongly understood number (e.g., 29 → 92). Interestingly, evaluation of the direction of the estimation error also corroborated the notion of German-speaking children being more prone to mix up tens and units: German-speaking children systematically under-/overestimated the position of a given number in accordance to whether mixing up tens and units would bias an under- or overestimation (i.e., 72 → 27 vs. 28 → 82, respectively). Although place-value understanding influenced estimation performance in both cultures in a very similar way, German-speaking children were at a particular disadvantage due to the fact that place-value information conveyed by their number word system does not correspond to the ordering of tens and units when expressed symbolically.

Study 4 was intended to evaluate the importance of children's early place-value knowledge on their later arithmetic performance in a longitudinal approach. So far, most of the evidence suggesting basic numerical competencies to be a building block for later development of numerical / arithmetical skills and capabilities comes from cross-sectional studies (e.g., Holloway & Ansari, 2009). In the present study it was investigated whether early place-value knowledge as operationalized by children's performance in a transcoding as well as a magnitude comparison task in first grade is a reliable predictor of performance in an addition verification task in third grade. Particular interest was paid to the influence of more specific predictors reflecting processes of place-value integration in basic number processing (e.g., the number of inversion errors in transcoding or the size of the compatibility effect in magnitude comparison) on comparable effects in basic arithmetic three years later (e.g., the carry effect). The results indicated that performance in basic numerical tasks served as reliable predictors of later arithmetic performance. Apart from this rather general influence of basic numerical knowledge on later performance we also observed evidence for a specific influence of children's early place-value knowledge in first grade on addition performance in third grade: Children with a more elaborate understanding of the place-value structure in grade 1 showed better performance in the arithmetic task three years later. More particularly, these children committed less place-value related errors in carry addition problems, thereby exhibiting a comparably smaller carry effect. These findings suggest an important role of early place-value understanding for the later development of numerical / arithmetical competencies (see also Study 2 for influences of the place-value structure of the Arabic number system on adult addition performance).

Section 4: The neuro-cognitive underpinnings of place-value integration

In ***Study 5***, the neural correlates of place-value integration were investigated by functional MRI. In a previous study Nuerk, Geppert, van Herten, and Willmes (2002) found that in a verification version of the number bisection task triplets crossing a decade boundary were responded to significantly slower than triplets that stayed within the same decade (e.g., 25_28_31 vs. 23_26_29). The authors interpreted this as indicating the need for additional processing of place-value information to evaluate the distance between e.g., the second and the third number when crossing a decade boundary. The interest of the current study was to examine whether there is a specific neural correlate of place-value processing, implicating a cortex area with higher neural activity while processing triplets crossing into the next decade compared to triplets with no decade crossing. As previously observed by Wood, Nuerk, and Willmes (2006, see also Knops, Nuerk, Sparing, Foltys, & Willmes, 2006), processes of unit-decade integration were associated with more pronounced activation in the IPS, a cortex location affiliated with the processing of magnitude information in general. However, as overall range (i.e., the distance between the two outer numbers) was accounted for in the parametric analyses for triplets with and without a decade crossing this activation dissimilarity cannot be explained by differences in holistic magnitude measures. Instead, the data argue for specific processing demands associated with the required processing of place-value information for triplets crossing into the next decade meaning that place-value information is not only retained in the representation of two-digit number magnitude but also explicitly processed when participants encounter two-digit numbers. Thereby, these results indicate a specific neural correlate of place-value processing / integration.

Study 6 examined the influence of the disruption of the mental number in neglect patients – as first reported by Zorzi, Priftis, and Umiltà (2002) – on the processing of place-value information. Zorzi and co-workers (2002; see also Zorzi, Priftis, Meneghello, Marenzi, & Umiltà, 2006) observed that, when asked to indicate the midpoint of a numerical interval, patients' perceived midpoints of the intervals were misplaced to the right (reporting 7 as being midway between 1 and 9) similarly to these patients impairments in line bisection. In the present study, interest was paid to the question whether neglect-related deficiencies also affect the processing of place-value information as conceptualized by decade crossings in a verification version of the number bisection task (e.g., 23_26_29 vs. 25_28_31; see also Study 5). Results showed that neglect patients' performance for triplets crossing a decade boundary was significantly worse than performance of the control group. Furthermore, neglect patients were particularly impaired when the decade crossing occurred between the first and the second number of the triplet rather than when the decade crossing occurred between the second and the third number (e.g., 28_31_34 vs. 25_28_31). These results denote that not only the processing of number magnitude information but also the representation of the place-value structure of the Arabic number system is impaired in neglect patients.

Section 5: A computational model of place-value integration in two-digit number comparison

Finally, ***Study 7*** provided first computational evidence helping to differentiate between three different representational formats of two-digit numbers (i.e., holistic, strictly decomposed, or hybrid). Particular attention was paid to the dissociation of the strictly decomposed and the hybrid model as there is currently no empirical data that allows distinguishing between these two assumptions. To pursue these issues three computational models were programmed realizing an either holistic, decomposed or hybrid representation of two-digit number magnitude. In both the comparison of a given number to a fixed as well as a variable standard the strictly decomposed model produced the data that fitted the empirical data best. However, not only did the decomposed model produce the most realistic data pattern, its RT estimates also reflected specific characteristics previously assumed to be determined by the holistic part of the hybrid model. In this context, regression analysis showed that a predictor representing a measure of (logarithmic) overall distance between the to-be-compared numbers accounted for a reliable part of the variance even though there is no unitary representation of overall distance in the strictly decomposed model. Interestingly, the inclusion of such a predictor was one reason for Nuerk, Weger, & Willmes (2001; Nuerk and Willmes, 2005) to propose a holistic number magnitude representation apart from representations coding the magnitudes of its constituting digits. Taken together, analysing the computational model data implies that the standard effects of numerical cognition on symbolic digital input can be accounted for by a model proposing the strictly decomposed representation of units, tens, hundreds, etc.

After having summarized the most important aspects of the individual studies presented in the current thesis, these so far separate lines of evidence shall be consolidated and discussed in the General Discussion section.

Section 7

General Discussion

GENERAL DISCUSSION

In the last section of this thesis a general discussion integrating the results of the individual studies shall be given. This discussion will be subdivided into two main paragraphs. First, in the *synopsis* the major results of the current thesis will be evaluated as regards their relevance for corroborating the notion of a generalizable influence of place-value properties (as reflected by evidence for decomposed processing of tens and units) on different aspects of numerical cognition, such as the nature of the mental number line and its development. Subsequently, the results concerning individual and cultural differences will be discussed. Second, the specific findings of the individual studies going beyond the scope of the initial synopsis will be discussed in the *discussion and evaluation* section following the general outline of this thesis differentiating evidence on representational, developmental, neuropsychological, and computational aspects of the influence of the place-value structure of the Arabic number system on human numerical cognition. Finally, the present thesis will be closed by a *conclusions* section reiterating the most important findings.

SYNOPSIS

At the end of the general introduction, the main aim of the current thesis was identified as trying to provide a comprehensive evaluation of the influences of the place-value structure of the Arabic number system on numerical cognition by a) further establishing the generalizability of decomposed processing of two-digit number magnitude and b) an investigation of possible individual and cultural differences. The results of the individual studies were meaningful to both of these topics and will be discussed accordingly.

The generalizability of decomposed processing of multi-digit numbers

Considering the generalizability of decomposed number magnitude processing, it was sought to extend existing evidence on four different aspects: (i) the nature of the mental number line, (ii) the subject of the numerical tasks for which evidence for decomposed processing was obtained, (iii) the longitudinal developmental influences of place-value understanding, and (iv) processing specificities of decomposed number magnitude processing – each of which will be elaborated on in turn.

The nature of the mental number line

Concerning the nature of the mental number line the results of Study 1 observing a reliable unit-decade compatibility effect for comparisons to a fixed internal (memorized) standard is hard to reconcile with the assumption of the internal number magnitude representation being holistic in nature. Instead, the present findings suggest that the mental number line may also involve representations of the structural characteristics of the Arabic number system such as its base-10 place-value properties. This latter interpretation is further corroborated by the results of the patient study (cf. Study 6). Hemispatial neglect not only affected participants' performance in a number bisection task regarding analogue scaling properties but also the processing of place-value information (i.e., increased difficulty when a to-be-evaluated triplet crossed a decade boundary). These findings again indicated that the mental number line incorporates structural information on the place-value organization of the Arabic number system – or as Nuerk, Weger, and Willmes (2001) put it, the existence of "decade breaks in the mental number line (p. B25).

Further evidence for decomposed processing of tens and units

The results of Studies 1 and 6 were based on tasks for which there is already evidence suggesting decomposed processing of number magnitude, i.e., the number magnitude comparison task (e.g., Nuerk et al., 2001; Ratinckx, Nuerk, van Dijk, Willmes, 2006; Moeller, Fischer, Nuerk, & Willmes, 2009a) and the number bisection task (see Nuerk, Geppert, van Herten, & Willmes 2002, Moeller, 2006). However, the present thesis also aimed at gaining further evidence for the importance of structural place-value information from tasks previously not primarily considered to be sensitive to place-value related influences. At the level of basic arithmetic operations, i.e., addition, it was observed that the need of a carry operation (requiring specific processes of place-value integration, e.g., 36 + 48) specifically increased processing demands on the unit digits of the summands. Moreover, as observed in Study 4 of the current thesis it seems that children's performance in number line estimation tasks is accounted for best by a model suggesting two separate representations of single- and two-digit numbers. Thereby, the results again indicated that place-value attributes of the Arabic number system are retained in the human representation of numerical magnitude.

Longitudinal developmental aspects of place-value understanding

The preceding paragraphs were aimed at establishing decomposed processing of multi-digit numbers as a general processing principle. Additionally, first evidence for influences of children's place-value knowledge on numerical development was described (cf. Study 4).

However, unlike in previous studies (e.g., Holloway & Ansari, 2009) the use of two-digit stimuli allowed for the evaluation of place-value processing. In Study 3 of the current thesis it was observed that the worse children's place-value understanding was in first grade the worse they performaned in an addition task administered in third grade. Notably, a higher number of inversion transcoding errors in first grade was associated with a larger number of errors in carry addition problems and a larger carry effect in third grade, both indicating impaired processes of place-value integration. Thus, these data not only suggested a reliable influence of place-value understanding on numerical development but also indicated rather direct developmental trajectories for the representation of the place-value structure of the Arabic number system.

Specificities in processing place-value information

Further indication for the validity of decomposed processing to be a rather general characteristic of multi-digit number processing comes from modelling results. On the one hand a two-linear model suggesting two separate representations of single- and two-digit numbers fitted children's estimation pattern in the number line task reliably better than did a holistic logarithmic model (see Moeller, Pixner, Kaufmann, & Nuerk, 2009b; Study 3). On the other hand, directly contrasting all three proposed models of two-digit number magnitude representation (i.e., holistic, strictly decomposed, and hybrid) using computational modelling revealed that the compatibility effect could not be simulated by the holistic model but by the strictly decomposed and the hybrid model. Taking into account model proximity it was found that all empirical effects observed for symbolic two-digit number comparison could be simulated by a strictly decomposed model. Against this background the question arises whether there is a specific neural correlate associated with the processing of place-value properties. Assessed by the number bisection task, increased activation in posterior parts of the IPS was found for the evaluation of triplets crossing a decade boundary that required specific processing of place-value information (cf. Study 5). This again is in line with the interpretation suggesting the mental number line to not only represent analogue magnitude information but to also involve information about place-value properties.

Taken together, the results of the current thesis provide further evidence for decomposed processing of multi-digit numbers (reflecting a crucial role of the place-value structure of the Arabic number system) to be a general processing principle rather than a specificity to particular presentational formats, tasks, and/or participant populations. In turn, this strongly suggests that place-value information is an inherent part of the mental representation of number magnitude (possibly along a mental number line) as it seems to be retained in numerical cognition in general.

Future perspectives on the generalizability of place-value processing

On a more abstract level, the assumption of place-value structuring as a general processing characteristic raises important questions regarding the overall validity of previous findings in number processing research. In particular, when acknowledging the results of the computational modelling study further research is needed to investigate the full range of consequences of the fact that typical empirical effects of two-digit number comparison were accounted for best by a strictly decomposed model, not incorporating any holistic magnitude representation at all. Based on these findings, is it justified to operationalize bisection range in a number bisection task in terms of the overall distance between the two numbers defining the to-be-bisected interval (e.g., Nuerk et al., 2002; Study 5; Göbel, Calabria, Farnè, & Rossetti, 2006; Loftus, Nicholls, Mattingly, & Bradshaw, 2008), or the distance between the correct middle number and a given probe by giving the overall distance between the probe and the correct middle number (e.g., Nuerk et al., 2002; Study 5; Loftus et al., 2008)? The current data suggest that it might be more appropriate to distinguish between the decade and the unit distance. And indeed, such a distinction may be useful. Consider the latter case of distance to the middle number when the following interval should be bisected e.g., 29_52_75. In such an instance rejecting an incorrect middle number should be harder when the probe differs at the unit digit only as when it differs at the tens position as well; even when overall difference were the same (e.g., 29_56_75 vs. 29_48_75, respectively; please note that absolute overall distance to the middle is 4 for both probes). Thus, it may be more appropriate to operatinalize the distance between the correct middle number and the probe in terms of its distance at the tens as well as the unit position rather than computing an overall distance measure (i.e., 29_56_75 → decade distance: 0, unit distance 4; 29_48_75 → decade distance: 1, unit distance: 6). Similarly, in addition verification (cf. Study 2) it might be the case that for the problem 18 + 24 = 32 the probe 35 may be harder to reject as incorrect when compared to the probe 29 since it shares the correct tens digit even though overall split between both probes and the correct result is identical (i.e., 3 in this example). Again, the results of the current thesis imply that it should be more suitable to account for the split in terms of decade and unit distance as compared to an overall distance measure not accounting for place-value related differences of the probes. In light of the longitudinal results of Study 4 the latter distinction becomes even more relevant as it would provide an additional way of analyzing place-value influences not only for the correctly solved problems (e.g., 18 + 24 = 32, requiring a "yes" response) but also for those solved incorrectly (e.g., 18 + 24 = 35, requiring a "no" response; actually representing 50% of all items). So far, incorrect probes have been chosen to differ at either the tens or the unit position but did not discern between above introduced decade-consistent and decade-inconsistent probes. However, Domahs, Nuerk, and Delazer (2006; Domahs et al., 2007) were already able to show that such kind of decade consistency had a reliable effect on performance in simple multiplication: participants produced more decade-consistent than inconsistent faulty responses and it took them longer to reject a decade-consistent as compared to a decade-inconsistent probe (see introduction for a more elaborate explanation of the decade consistency effect in multiplication). Considering this recent evidence on place-value influences in mental arithmetic, closer investigation differentiating effects of decade and unit distance seems a promising approach to get further insights into the generalizability of the influence of the place-value structure of the Arabic number system on numerical cognition.

Anyway, apart from the results suggesting place-value related influences to reflect a comprehensive processing principle rather than specific effects associated with a particular task, the current thesis also provided evidence suggesting that place-value understanding is nevertheless subject to individual and cultural differences. Aspects concerning inter-individual as well as inter-cultural variation in processing place-value information will be addressed in the following paragraph.

Individual and cultural differences in processing place-value information

Even when accepting the assumption that representations of the Arabic number system's place-value structure are vitally involved in numerical cognition, the question of possible inter-individual and inter-cultural differences remains.

As regards the latter, Miura and colleagues (Miura & Okamoto, 1989; Miura et al., 1994) showed that the transparency of the number word system of a given language concerning the way in which units, tens, etc. are verbally recoded is an important determinant of children's mastery of the place-value structure. Against this background Study 3 of the current thesis was conducted to evaluate the influence of such language differences of number word systems in a totally non-verbal task, the number line estimation task. Despite the fact that performance of both German- and Italian speaking children was accounted for best by a two-linear model with a fixed breakpoint at 10 (see above for a discussion), German- and Italian-speaking children differed reliably in the accuracy of their estimates upon the hypothetical number line. Compared to their Italian-speaking counterparts, not only were estimates of German-speaking children less accurate in general, they also specifically under-/overestimated the position of numbers for which confusing tens and units would result in a relatively smaller/larger number as (i.e., marking the position of 27 when asked to estimate the position of 72). This finding suggested an important influence of place-value properties in numerical cognition (superiority of the two-linear over the logarithmic fitting) as well as a moderation of such influences by cultural differences (such as the structure of the number word system) even for performance in a completely non-verbal task

From this a number of questions arise concerning possible language related effects (possibly determined by differences in the respective number word systems) on other numerical tasks. So far, language differences have been observed and related to differences in the verbal coding of place-value information only for the case of number magnitude comparison (see Nuerk, Weger, & Willmes, 2005, Macizo & Herrerá, in press for adult; Pixner, 2009 for children data; see also Miura & Okamoto, 1989; Miura et al., 1994 for the explicit naming of tens and units; Domahs et al., in press, for the case of finger counting). When singling out the larger of two numbers presented in digital notation the regular compatibility effect was reliably smaller but still significant for English-speaking participants as compared to their German-speaking counterparts. Nevertheless, for the case of two-digit number words, which follow the decade unit structure of the digital format, English-speaking participants showed a marginally significant reversed compatibility effect (see also Macizo, Herrerá, Paolieri, & Román, in press for a reversed compatibility effect for non-inverted Italian number words).

Thus, the present results for the number line estimation task suggested that there may be even more differences related to the processing of place-value information. For instance, it may be hypothesized that the carry effect in addition may be smaller for participants speaking a language without an inversion of the order of tens and units in two-digit number words. As verbal recoding of tens and units does not interfere with the order of tens and units in symbolic notation executing the required unit-decade integrations due to carrying the decade digit of the unit sum might be less time consuming and/or less susceptible to errors. So, it is up to future studies to investigate whether differences in place-value coding between different number word systems has consequences going even beyond the effects reported in this thesis. Against the background of the accumulating evidence for the importance of place-value representations such effects should be observable – at least in children whose processing of numerical (including place-value) information is less automated.

With respect to individual differences, the results of Study 4 indicate that there are indeed differences in the capability to process place-value information. Also, such differences in early place-value understanding are predictive not only for later arithmetic performance in general but also for place-value related aspects of later arithmetic, in particular. Please note that this is the first time that the longitudinal influence of children's early place-value understanding was investigated. The current data suggest that individual differences in place-value mastery exist and seem to be relatively stable over time – thereby, implying an important but so far largely under-investigated role of place-value understanding in numerical development.

Taken together, Studies 3 and 4 of the current thesis provided evidence for both individual and cultural differences regarding the mastery of the place-value structure of the Arabic number system. This is important for at least two reasons: First, it strongly argues for place-value understanding to be considered a specific numerical representation of its own, rather than being implicitly subsumed in the representation of number magnitude. Second, it raises the question whether these observed individual / cultural differences may be of diagnostic and / or predictive value in the way comparable differences for other numerical representation have already been employed (see e.g., Delazer et al., 2006; Holloway &

Ansari, 2009; Halberda, Mazzocco, & Feigenson, 2009 for the case of number magnitude representation). The current data support such a view since children's early place-value understanding serves as a reliable predictor of their later numerical / arithmetical performance (see Study 4) and also accounts for language differences concerning the transparency of number word systems.

Acknowledging its reliable influence on numerical development a representation of the place-value structure of the Arabic number system should be incorporated in any model capturing the development of numerical / arithmetical competencies in children. In the following paragraph different models of numerical development shall be evaluated regarding their comprehension of place-value representations.

Place-value representation in a current model of numerical development

As already described in the general introduction the dominant structuring principle of the Arabic number system apart from its ciphered nature (see Chrisomalis, 2004 and introduction for details) is the principle of place-value organization. Additionally, the results obtained in the individual studies of the current thesis argue for an important role of place-value information in the representation of multi-digit numbers as well as for basic arithmetic involving such numbers. Nevertheless, to the best of our knowledge there is currently no (neuro)cognitive model of numerical development explicitly incorporating place-value understanding. For instance, von Aster and Shalev (2007) proposed a four-staged developmental model of numerical cognition in which they associated specific numerical/calculation competencies to cognitive representations underlying these capabilities on the one hand, and to brain areas subserving these abilities. Place-value understanding is not explicitly conceptualized within this model as well. Yet, it seems plausible to assume that place-value understanding comes into play at stages three and/or four (i.e., Arabic number system and the ordinality of the mental number line, respectively) of the von Aster and Shalev (2007) model as it is a general ordering principle of the Arabic number systems and was also shown to influence the development of a spatial mental number line representation (cf. Moeller et al., 2009b; Study 3). This interpretation is corroborated by the authors' description of typical impairments of children with dyscalculia for each of the four developmental stages. For stage three, transcoding errors related to the inversion property of e.g., the German number word system are mentioned, clearly tapping place-value processing. Nevertheless, the authors did not specify the issue of place-value representations within the context of their model of number acquisition. Moreover, even when considering other theoretical conceptualizations of numerical development and developmental dyscalculia, in particular, the influence of e.g., insufficient place-value understanding is widely neglected. In line with Karmiloff-Smith's developmental theory (1996; 1998), Rubinsten and Henik (2009) suggested different underlying deficits and their interactions to determine developmental dyscalculia and/or mathematical learning disorders. Amongst others, the authors proposed deficits in magnitude/quantity processing, working memory, attentional resources, and executive functioning to be relevant for mathematics impairments, but again did not specifically address processing of place-value information.

Taken together, there is so far no comprehensive theory of numerical development, either typical or atypical, explicitly considering place-value understanding to be a relevant and necessary step in emerging numerical cognition. Taking into account the finding obtained in the current thesis, in particular the longitudinal influence of early place-value understanding on later arithmetic performance it would be desirable to further conceptualize the developmental impact of place-value knowledge by explicitly incorporating it into a model of numerical development such as the model by von Aster and Shalev (2007). Within this framework, implementing a representation of the Place-value structure of the Arabic number system would not be too problematic. As von Aster and Shalev (2007) propose children get confronted with the issue of place-value coding first as they experience number words above 10 (or 20 when acknowledging the special character of teen numbers). Therefore, the postulated representation of place-value information should be triggered in step 2 of the von Aster and Shalev (2007) model. However, as the authors suggest for the role of working memory (dotted arrow in Figure 1) it may be the case that a place-value representation develops not in an all-or-nothing manner but gets more and more established and relevant over the four developmental steps: For representations of concrete quantity as proposed in step 1 no place-value representations should be acquired. However, with the experience of number words within and/or above the teen range in step 2 place-value issues become more and more important to build up a successful connection between the respective number word and its quantitative meaning. In step 3, the role of place-value information is even more prominent as within the to-be-learned Arabic number system it reflects one of the core organization principles (together with ciphering, see Chrisomalis, 2004). Thus, to correctly reflect a given number in symbolic Arabic notation place-value coding has to be understood (see Zuber, Pixner, Moeller, & Nuerk, 2009). Yet, the development of place-value representations may still be in progress in step 4 when the spatial mental number line representation is established. The results of Study 3 clearly show that even non-verbal representations of spatial number magnitude representations are influence by place-value properties (see also Moeller et al., 2009b). Additionally, the longitudinal influences of early place-value understanding on arithmetic performance (cf. Study 4) suggest place-value influences to actually generalize to children's arithmetic thinking. To sum up, a representation of the place-value structure of the Arabic number system within the model of von Aster and Shalev (2007) may be conceptualized similarly to the influence of working memory (see Figure 1, dotted arrow) as an arrow starting in step 2 and ascending until step 4 (see Figure 1, broken arrow and shaded area below) representing both the development of the place-value representation itself as well as its increasing importance for the numerical capabilities associated with the consecutive developmental steps.

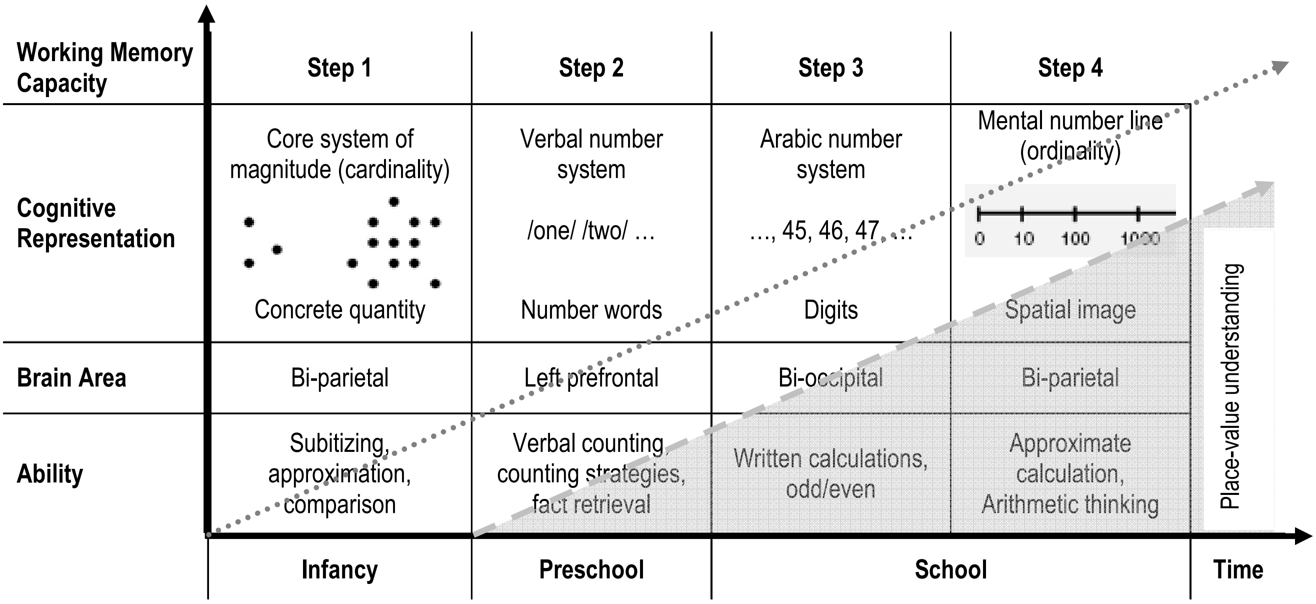


Figure 1: *Model of numerical development by von Aster and Shalev (2007, p. 870) adapted and extended by a representation of the place-value structure of the Arabic number system (broken arrow and shaded area below)*

On a closer inspection of the quantity representations proposed for Steps 1, 2, and 3 it has to be noted that only from step 2 on symbolic coding (verbal and digital) is assumed while step 1 quantity representations rely on non-symbolic coding of concrete quantities. This distinction seems to be motivated by observations suggesting that even preverbal infants are able to discriminate between numerical quantities (e.g., Feigenson, Carey, & Hauser, 2002; Feigenson, Dehaene, & Spelke, 2004; Xu, Spelke, & Goddard, 2005; for a review see Cordes, & Brannon, 2008). The place-value structure of the Arabic number system first comes into play with the development of symbolic quantity representations in step 2 when tens and units have to be discerned when confronted with the respective number words. Decomposed processing of tens and units can be assumed to start at this developmental stage as well and to get more and more automated with increasing age and experience (cf. Pixner, Moeller, Zuber, & Nuerk, 2009) as it allows for representing any given number in a very efficient way by a limited set of symbols (cf. Verguts & de Moor, 2005). The results of Studies 3 and 4 of the present thesis indicated that apart form inter-individual differences the development of decomposed processing of multi-digit numbers is also liable to inter-cultural differences. The structure of different number word systems may either hinder or corroborate the successful mastery of the place-value structure of the Arabic number system. Synced with the results of Study 7, in which the strictly decomposed model was found to account best for typical empirical effects in two-digit number comparison model, this may have important implications for the nature of the human representation of number magnitude. In particular, the fact that a holistic representation of two-digit number magnitude is not mandatory to simulate typical empirical effects in number comparison suggests that the decomposed representations of e.g., tens and unit do not seem to be integrated into one entity representing the overall magnitude of a given number, at least for symbolic notation. In turn, the current findings imply that at most the representations of single digit numbers up to 9 may be grounded on a holistic non-symbolic counterpart whereas representations of multi-digit numbers above 10 may involve no direct reference to a non-symbolic quantity representation. Please note that this interpretation fits nicely into the ongoing debate on notation dependent or notation independent / abstract representation of number magnitude (e.g., Cohen-Kadosh & Walsh, 2009; Piazza, Pinel, Le Bihan, & Dehaene, 2007; Cohen-Kadosh, Cohen-Kadosh, Kaas, Henik, & Goebel, 2007).

PRÉCIS

All in all, the results of the this thesis, obtained from a broad variety of numerical tasks administered to different populations (i.e., German- and Italian-speaking children, healthy adults, and neglect patients) and assessed using different experimental methodologies (i.e., reaction time, eye-tracking , and fMRI paradigms) lend further support to the notion that a mental representation of the place-value structure of the Arabic number system exists.

Furthermore, converging evidence is provided that the latter exhibits a task invariant influence on human numerical cognition and its development. Evidence on such a general validity of place-value influences comes from studies indicating (i) that the mental number line is not analogue holistic in nature but involves place-value information (cf. Studies 1, 5, and 6), (ii) that indicators of decomposed processing of two-digit numbers were observed from new basic numerical as well as basic arithmetical tasks [i.e., number line estimation (cf. Study 3) and addition (cf. Study 2), respectively], (iii) that decomposed processing seems to develop culture invariant but is nevertheless influenced by attributes of the respective language's number word system (cf. Study 4), (iv) that models implying decomposed processing outperformed models of holistic processing in direct comparisons (cf. Studies 3 and 7), and (v) that specific neural correlates of processing place-value information can be identified. Hence, the results of this thesis can be summarized as suggesting decomposed processing of multi-digit numbers along place-value constraints to be a comprehensive, culture invariant, and developmentally important processing characteristic of human numerical cognition.

Apart from this integrative view on the implications of the current thesis going beyond the scope of the individual studies, the results of each individual study are nevertheless important on its own. Therefore, the following paragraphs address these results in a less integrative manner, instead focussing on the scientific contribution of each study on a more individual level.

EVALUATION AND DISCUSSION

In the following section, the results of the previously presented studies will be discussed in respect to two major topics: First, it will be evaluated in what way behavioural, neuropsychological and computational evidence provided by the individual studies corroborated the understanding of how the place-value structure of the Arabic number system is reflected in the human number magnitude representation. In a second paragraph, the influence of place-value understanding on the development of numerical competencies will be elaborated on.

On the representation of the place-value structure of the Arabic number system

Addressing the question whether and if so, in what way the place-value structure of the Arabic number system is retained in our representation of two-digit numbers, the results of Studies 1, 2, 5, 6, and 7 are meaningful regarding (i) the influence of presentational format, (ii) place-value processing in more complex numerical tasks such as the number bisection task and mental addition, and (iii) the nature of two-digit number magnitude representation (i.e., holistic vs. decomposed vs. hybrid). These facets will be discussed in turn.

Presentational format and place-value processing

Recent criticism suggested that decomposed processing of tens and units may be determined by either presentational format or mode of presentation. In 2005, Zhang and Wang claimed decomposed processing of tens and units to be limited to the case when both to-be-compared numbers (i.e., probe and standard) are presented and thus represented externally – an argument also put forward by Ganor-Stern, Pinhas, and Tzelgov (2008). On the other hand, these authors presented evidence suggesting holistic processing of two-digit numbers in which the place-value structure of the Arabic number system is not retained when (i) a given number has to be compared to a fixed internally memorized standard (cf. Zhang & Wang, 2005) and when (ii) the to-be-compared numbers are presented sequentially instead of simultaneously (cf. Ganor-Stern et al., 2008). In this context, the results of Study 1 clearly indicate that decomposed processing did not depend on an external and simultaneous presentation of both probe and standard. Rather, we were able to show that the lack of unit based effects (such as the compatibility effects) in studies using a fixed internal standard may be attributed to methodological problems. In most of these studies (e.g., Zhang & Wang, 2005; Dehaene, Dupoux, & Mehler, 1990) a fixed standard ending on five such as 55 or 65 was used. Thereby, unit distance was limited to rather small values with a maximum of 4 (e.g., 55_79). However, already Nuerk et al. (2001) observed that the compatibility effect was more pronounced for large unit distances (i.e., 4 - 8) and weaker or even not significant for small unit distances (i.e., 1 - 4). Thus, it is not very surprising that unit-based effects such as the compatibility effect have not been observed in studies using a fixed standard ending on 5 (see also Nuerk & Willmes, 2005 as well as Study 1 for a more detailed discussion of this point). However, by the time Study 1 was conducted the paper by Ganor-Stern and colleagues (2008) was not yet published so the issue of simultaneous vs. sequential presentation was not tackled in this study but should nevertheless be discussed.

Sequential vs. parallel presentation of to-be-compared numbers

Ganor-Stern et al., (2008) observed a regular compatibility effect only when the to-be-compared numbers were presented simultaneously but not when they were presented sequentially. From this finding the authors concluded that the internal representation of tow-digit number magnitude representation is holistic in nature rather than decomposed. However, there are a number of theoretical and methodological problems concerning the validity of the sequential presentation condition as used by Ganor-Stern et al. (2008). From a theoretical point of view it may not be entirely appropriate to directly compare simultaneous and sequential processing conditions without referring to the problem of processing stages involved in either of the two conditions. Already back in 1977 Banks argued in his influencing paper on semantic processing in comparative judgements that differences between simultaneous and sequential presentation conditions may be determined by the direction of what he called relative coding of the to-be-compared numbers. In the sequential condition participants may have build up expectations about the first number being the large or smaller one even before having seen the second number. Thereby, when the second number is presented this expectation is either confirmed or disconfirmed. In the first case, no further processing of the two numbers magnitude is required and only in the latter case additional evaluations of the magnitude relations between tens and units may be initiated. Thereby, the results for simultaneous and sequential presentation conditions may stem from two completely different underlying processes and thus direct comparisons should be treated with caution.

Moreover, the issue of confirming / disconfirming expectations about the second number may be even aggravated by the realization of the sequential presentation condition. From a methodological point of view it seems problematic that to-be-compared numbers were presented sequentially at exactly the same position and with an inter-number-interval of only 200 ms without using a backward mask. Even when about one-third within-decade trials such as 43_49 were included to prevent participants from focusing on the tens position only (see Experiment 2 of Ganor-Stern et al., 2008), at this speed the task is more like a visual discrimination task on the tens position than a number magnitude comparison task requiring the processing of both tens and units. Imagine the case of a within-decade trial, with an inter-number-interval of just 200 ms and both numbers being presented at exactly the same position such an item is identified immediately by a change of the visual input at the unit position only. On the other hand, between-decade trials can be identified 100 % correct by a change of visual stimulus properties at the tens position. Thereby, the sequential condition as realized by Ganor-Stern et al. (2008) includes visual cues coding which digit is relevant for which trial that are 100 % valid. Based on this methodological flaw synced with above considerations on possible representational differences between sequential and simultaneous presentation conditions the claim of a holistic magnitude representation engaged in the sequential processing condition seems much too strong. A future study preventing visual cues to be 100 % valid by e.g., employing a backward mask presented after each individual number or jittering the positions of the sequentially presented numbers may indeed provide evidence for a decomposed processing of tens and units even when the to-be-compared numbers are presented sequentially. Finally, Wood, Mahr, and Nuerk (2005) showed that when the to-be-compared numbers are presented one after the other (with a 100 ms delay) the individual digits of the number presented first are compared by their magnitude. The authors also found that these intra-number comparisons worked against the overall compatibility effect. Again, this questions the conclusions drawn by Ganor-Stern et al. (2008).

Taking into account above arguments as well as the results of Study 1 it seems reasonable to assume that the internal representation of number magnitude may not be holistic as claimed by Zhang and Wang (2005) and Ganor-Stern et al. (2008; see also Zhou, Chen, Chen, & Dong, 2008) but is decomposed as well, thereby retaining the place-value structure of the Arabic number system.

The latter proposition that the place-value structure of the Arabic number system is an integral constituent of the human representation of number magnitude is corroborated by the variety of tasks for which an influence of place-value knowledge on performance was observed in the current thesis. Apart from the very basic number comparison task further evidence for the importance of processes of place-value integration come from the number bisection task, mental addition as well as a number line task which shall be discussed in the following.

Further evidence for place-value influences from more complex tasks and mental arithmetic

Place-value information in the number bisection task

In the number bisection task triplets involving a decade crossing (e.g., 23_27_31) were more difficult to evaluate than triplets staying within the same decade (e.g., 21_25_29; Nuerk et al., 2002; see also the following discussion on Study 5 below). This is particularly remarkable as this specific item property is not primarily important for solving the task. Instead, the number bisection task was originally assumed to rely on a default solution strategy of magnitude manipulations upon the holistic mental number line, only (e.g., Dehaene & Cohen, 1997; Cohen & Dehaene, 2000). Building on this considerations of performance in the number bisection task to be driven primarily by analogue number magnitude representations (seeing each triplet as a cut-out part of the mental number line), thus, may indicate that place-value information is nevertheless represented by the mental number line. This interpretation is supported by the results of the patient study (Study 5). Not only did the neglect patients exhibit performance impairments associated with them neglecting the left part of the mental number line (cf. Zorzi, Priftis, & Umiltà, 2003) and thus the numerically left part of each triplet also for the whole range of two-digit numbers. More specifically, they had particular problems for decade crossing triplets. The decrease of performance accuracy due to a decade crossing was more pronounced for the neglect patients than for the participants of the control group. On a representational level this may suggest that the processes of integration place-value information are particularly impaired in hemispatial neglect. Furthermore, it was observed that the decade crossing effect for neglect patients interacted with the position of the decade crossing within the triplet. Neglect patients' performance was reliably worse in triplets in whom the decade crossing occurred between the first and the second number (e.g., 28_32_36) compared to the second and third number (e.g., 24_28_32). So, in addition to their impairment determined by neglecting the relatively smaller numbers within the triplets, processes of place-value, i.e., unit-decade, integration also deteriorated the further left they occurred in the neglected area upon the mental number line. This finding is of particular interest as the similar gradients of impaired magnitude as well as impaired place-value processing again suggested that place-value characteristics of numbers may be inherent in the human number magnitude representation.

Moreover, this evidence from neuropsychological patients was further corroborated by a functional MRI study presenting a similar number bisection task to healthy adult participants to evaluate the neural correlates of processes of place-value integration such as the decade crossing effect. As previously observed for the case of number magnitude comparison and in particular the compatibility effect reflecting demands on place-value processing (cf. Wood, Nuerk, & Willmes, 2006) increased activation of the intraparietal sulcus (IPS), bilaterally was associated with specific processing of decade crossings. Interestingly, the IPS is also generally agreed to be essentially involved in the representation of number magnitude in general (see Dehaene, Piazza, Pinel, & Cohen, 2003; Nieder & Dehaene, 2009, for reviews). And again, these data suggest that the place-value structuring principle of the Arabic number system may be implemented into the human representation of numerical magnitude. Nevertheless, triplets crossing into the next decade were also associated with increased neural activation in prefrontal cortex areas. Previously, increased activation of prefrontal sites was observed in mental addition in general and for carry addition problems / borrowing subtraction problems, in particular (e.g., Kong et al, 2005; Imbo, Vandierendonck, & De Rammelaere, 2007). Both, the carry as well as the borrowing procedure in addition and subtraction, respectively, require processes of place-value integration (e.g., updating the sum of the decade digits of the summand by the carry). Generally, this increased prefrontal activation is interpreted to reflect increased demands of cognitive control / working memory / or attentional processes in carry addition / borrowing subtraction problems required for keeping intermediate results in mind, monitoring the columnwise solution process and finally executing the carry by updating the decade digit of the result by the decade digit of the unit sum. Increased activation at similar sites associated with decade crossing triplets may thus implicate processes of unit decade integration to require executive / attentional as well as working memory resources supplementing the more demanding processing of magnitude information in these triplets. More particularly, this indicated that the structural organizing place-value principle of the Arabic number system requires specific processing. Thus, processing place-value information is not an inherent aspect of number magnitude representation but may involve different magnitude-based as well as more general cognitive processes. Moreover, which of these processes are actually performed or not may also depend on the task to be performed. While in number magnitude comparison tens and units seem to be encoded and processed in parallel (cf. Moeller et al., 2009a) performance in a more complex task such as the number bisection task was suggested to involve quite different basic numerical processes including bottom-up processing of specific stimulus characteristics as well as top-down mediated processes of e.g., plausibility evaluations (cf. Moeller, Fischer, Nuerk, & Willmes, 2009c).

Eye-tracking evidence on place-value processing in mental addition

As described currently, Moeller and colleagues (2009a; 2009c) were able to evaluate processing characteristics for numerical tasks. The authors chose the eye-tracking methodology to differentiate between different processing strategies (Moeller et al., 2009a) or the temporal specificities of different numerical processes (Moeller et al., 2009c). The evaluation of participants' eye fixation behaviour while engaged in a given task seems to be a valid tool to distinguish between different basic numerical processes because of two general assumptions about eye fixations: (i) According to the eye mind assumption (e.g., Rayner & Pollatsek, 1989) the fixation location of the eyes serves as a reliable indicator of what part of a stimulus is processed at the moment. And (ii) the immediacy assumption (e.g., Rayner &

Pollatsek, 1989) states that an object or stimulus is processed as long as it is fixated. Thereby, the number of fixations as well as their durations is a valid predictor of how important information from a specific part of a stimulus (e.g., a particular digit) is in the context of performing the task at hand.

Based on above considerations, the eye-tracking methodology should also offer the possibility to differentiate between different basic processes underlying place-value integration. A question pursued in Study 2 of the current thesis. Here, participants' eye fixation behaviour in an addition verification task was evaluated to investigate possible origins of difficulty in carry addition problems. We were able to identify two distinct processes associated with the increased difficulty of carry addition problems each reflecting processes of place-value integration. On the one hand, performing carry addition problems seemed to be associated with specific unit based calculations indicating the requirement of a carry when reaching or exceeding a certain level (i.e., a unit sum of ≥ 10). This suggested that particular processing of the unit digits of the summand was necessary whenever the sum of these digits becomes a two-digit number and thus the decade digit of the unit sum has to be carried to the decade position to update the decade digit of the result. On the other hand, the results indexed a process of carry execution to add to overall difficulty as well. In carry addition problems the decade digit of the result required specific processing, possibly being associated by the updating of the sum of the decade digits of the summands by the carry from the sum of the unit digits. Thus, being able to dissociate these different processes and to associate them with the processing of specific digits (i.e., units of the summands vs. tens of the result) again argues for a decomposed representation of number magnitude. When assuming the magnitude of a two-digit number to be represented as an integrated entity not retaining any place-value information there should performance in two-digit addition should not be determined by processes that can certainly be attributed to the processing of either tens or units. Thereby, the results of Study 2 not only corroborated earlier findings suggesting a prominent role of the unit digits of the summands in mental addition (Klein et al., under revision). Rather, the data also served as a generalization of the findings by Moeller and co-workers (2009a) as they provide further eye tracking evidence validating the assumption of parallel but decomposed processing of two-digit numbers not only in a task as basic as number magnitude comparison but also in mental addition.

So far, the evidence provided by the current thesis regarding the role of the place-value structure of the Arabic number system within the human representation of multi-digit number magnitude was empirical in nature. However, in Study 7 the question in which way two-digit number magnitude is represented (i.e., either holistically thereby not retaining the place-value structuring or decomposed or hybrid and thus referencing place-value principles) was pursued by a computational approach.

Computational evidence on the nature of two-digit number magnitude representation

On a more theoretical level the current thesis also added significantly to our understanding of how two-digit numbers are represented. As already elaborated on above, the empirical findings of the present thesis consistently provide evidence for the notion of two-digit number magnitude being represented and processed in a decomposed manner separating tens and units; and thus explicitly complying with the place-value structure of the Arabic number system. However, currently there are two theoretical frameworks capable of accounting for such decomposed processing: The notion of a strictly decomposed representation of multi-digit numbers separated into their constituting digits (cf. Verguts & De Moor, 2005) as well as the hybrid model of multi-digit number representation assuming an additional holistic representation of the overall magnitude of a number despite the separate representations of units, tens, hundreds, etc. (cf. Nuerk & Willmes, 2005; Nuerk et al., 2001). As there is currently no empirical paradigm allowing for differentiating between these two theoretical models, we addressed this question by a computational approach. Three computational models were programmed realizing (i) holistic (ii) strictly decomposed, and (iii) hybrid representations of number magnitude. When evaluating the descriptive adequacy of the data produced by each of these models it was found that the data by the strictly decomposed model fitted the empirical data best. In particular, Nuerk and colleagues (2001; see also Nuerk & Willmes, 2005) initially assumed a representation of overall number magnitude to accompany the separate representations of tens and units as in most regression analysis a measure coding the overall distance between the to-be-compared numbers was observed to be the most important predictor of task performance. However, assuming such an additional holistic representation, because a measure of overall distance between probe and standard was included as a reliable predictor in a regression analysis, was proved unnecessary by the current data. A measure of overall distance (i.e., difference between the logarithms) was found to be the most important predictor of item RT even for the data produced by the strictly decomposed model. Thereby, indicating that overall a strictly decomposed model of two-digit number representation accounted best for the empirical data. Please note that the superiority of the strictly decomposed model was evident even apart from the fact that the hybrid model is the least parsimonious model and should thus only be considered when controlling for the higher number of free parameters. Nevertheless, the computational data clearly argue for the human representation of two-digit number magnitude to comply with the power levels of the Arabic number system, thereby, retaining its place-value structure.

In summary, the results of the present thesis were meaningful in addressing the question whether and if so in what way the place-value structure of the Arabic number system is reflected in the human representation of two-digit numbers. Empirical evidence suggested that decomposed processing of tens and units is not driven by an external presentational format. Instead, the present data indicate that the internal representation of number magnitude is decomposed as well (cf. Study 1). Furthermore, this important initial finding was corroborated by a variety of observations indexing decomposed processing of tens and units in more complex numerical tasks such as the number bisection task and even locating a specific area within the intraparietal sulcus associated with processes of place-value integration (cf. Study 5). These imaging results were accompanied by patient data (cf. Study 6) again suggesting separate processing of tens and units as well as a representation of place-value information to be impaired by hemi-spatial neglect in a way previously assumed to be affect the spatial representation of number magnitude (i.e., the mental number line) only. Thereof, we inferred the mental number line representation of number magnitude to involve at least some kind of place-value information. In this vein, the results of Study 2 imply that place-value information is not only an important aspect / cue in basic numerical tasks but also specifically processed in basic arithmetic such as addition, in particular, when processes of place-value integration are required to calculate the correct result as in carry addition problems. Finally, these conclusions drawn on the current empirical data are backed by the computational results provided by a neural network model of two-digit number comparison. Here, it was found that a strictly decomposed model accounted best for the empirical data (cf. Study 7), thereby, indicating no need for an additional holistic representation as proposed by the hybrid model of two-digit number representation (cf. Nuerk & Willmes, 2005). Taken all these evidence together, an essential role of the place-value structuring principle of the Arabic number system for the adult human representation of number magnitude seems undisputed.

Nevertheless, recent findings from studies investigating the development of numerical competencies reported evidence suggesting an important role of place-value understanding even for early numerical development. The findings from the current thesis addressing the influence of the place-value structure of the Arabic number system on the development of numerical competencies as well as the developmental trajectories of place-value understanding will be discussed in the following section.

The developmental influence of place-value knowledge

In discussing the influence of place-value understanding on the development of numerical competencies in children two different lines of evidence pursued in the current thesis shall first be evaluated and then integrated. On the one hand, the place-value knowledge structure of the Arabic number system may influence the way basic numerical representations are shaped (cf. Zuber et al., 2009). This aspect of early place-value knowledge was addressed by Study 3. On the other hand, early place-value understanding may be an important predictor of later arithmetical skills as investigated in Study 4. Such a predictive association has already been observed for other basic numerical representations such as the representation of number magnitude (cf. Holloway & Ansari, 2009). These two aspects of the developmental influence of the place-value structure of the Arabic number system will be discussed in turn.

The influence of early place-value processing on the development of the mental number line

Only recently, Zuber and colleagues (2009) found that even a task as basic as transcoding a number from spoken to written format is influenced by aspects relating to place-value. The authors found that in German-speaking children about half of all transcoding errors committed were related to the inversion property inherent in German number words [e.g., writing down 72 instead of 27 when dictated “siebenundzwanzig” (literally: seven and twenty)] and thereby possibly determined by confusion in place-value integration of the single digits, i.e., difficulties in assigning each digit its corresponding power level as these are inverted in German number words as compared to symbolic notation. Thus, this indicated an influence of the place-value structure of the Arabic number system on other numerical representations such as the verbal representation of number words and related task such as transcoding. In this context, Moeller et al. (2009b) observed that place-value influences may not be limited to the development of transcoding skills but also to building up the spatial representation of number magnitude. In a number line estimation task Moeller and co-workers (2009b) found that children's estimates of the spatial position of a given number upon a hypothetical number line were accounted for best by a two-linear function with a break point at 10. From this data, Moeller et al. (2009b) concluded that children's early mental number line representation seems to differentiate between single- and two-digit numbers. In Study 3 of the present thesis we were able to show that this influence of the structuring principle of the Arabic number system exhibits its influence on the development of the spatial representation of number magnitude occurs not only in German-speaking children (cf. Moeller et al., 2009b) but also in Italian-speaking children, thereby suggesting it to be a rather general principle. Nevertheless, we also found evidence for a specific influence of the structure of the languages' number word systems. This was possible because unlike in German number words the order of tens and units in Italian number words corresponds to the left-to-right order of e.g., tens and units in symbolic notation. Base on this difference in the number word systems of German and Italian we hypothesized the estimates of German-speaking children to be less accurate as they were more prone to confuse tens and units because of their antidromic order in verbal and symbolic notation. The results confirmed our hypothesis: The estimates of Italian-speaking children were reliably more precise as those of their German-speaking counterparts. Even more interestingly, the finding that this advantage was specifically driven by the mislocalizing of items for which the confusion of tens and units leads to a large deviation between the correct position of the number and the actually marked position (e.g., 82 as compared to 65). Taken together, these results indicate that there is a rather general influence of the place-value structuring principle of the Arabic number system on the development of the spatial representation of number magnitude which is valid independent of cultural influences as long as the Arabic number system is used. However, the data also imply that this influence is moderated by the language specific attributes such as the number word system. When the order of tens and units in a language's number word system does not correspond to the order of tens and units in symbolic Arabic notation this is detrimental for children developing a precise and accurate spatial representation of number magnitude. On a broader level, this is interesting as an accurate spatial representation of number magnitude seems to be a reliable predictor of actual mathematics achievement as well as children's ability to solve unknown arithmetic problems (Booth & Siegler, 2008).

The influence of early place-value understanding on further numerical development

At this point and based on the observed influence of the place-value structure of the Arabic number system on the development of a spatial representation of number magnitude (cf. Moeller et al., 2009b; Study 3 of the present thesis) the question arises whether early place-value knowledge itself may serve as a building block for further numerical arithmetical development. This issue was pursued in Study 4 of the current thesis in a longitudinal approach. As already described above, results indicated that early place value knowledge in first class (as operationalized by inversion errors committed in transcoding as well as by the compatibility effect in a magnitude comparison task) reliably predicted performance in an addition task two years later in third grade. Thereby, Study 4 provided first direct and longitudinal evidence for early place-value knowledge to determine the future development of children's numerical / arithmetical competencies. Apart from this very important finding Study 4 may be considered innovative for another reason. The predictive value of early place value understanding on later arithmetic performance could be specified in an effect based approach. Therein, it is not overall performance in third grade addition that is predicted by overall performance in magnitude comparison and transcoding assessed in first grade. Instead, specific measures reflecting the accuracy of processes of place-value integration in addition (i.e., the number of errors in carry addition problems and the carry effect) were reliably predicted by inversion transcoding errors and the compatibility effect, again being a measure of processes of place-value integration. Thus, Study 4 suggested a persistent influence of place-value knowledge in numerical development, in particular, when evaluating specific effects related to the processing of place-value information at different stages of development and schooling. This is of specific interest as the results of Study 2 of the present thesis showed that influences of place-value processing are still observable in adult arithmetic (i.e., particular processing of the unit digits in two-digit addition when a carry is required). Thus, there is the possibility that difficulties in early place-value understanding may not only be a problem restricted to a certain time in numerical development. Instead, impaired place-value understanding at earlier stages of numerical development may determine later arithmetic performance even into adulthood. This assumption is further corroborated by the results of a recent eye-tracking study comparing the eye-fixation behaviour of children and adults in an addition verification paradigm (Moeller, Klein, & Nuerk, under revision). On the one hand, results indicated that there are considerable differences as regards the way children and adults select the correct probe. While children seemed to primarily rely on calculating the correct result, adults seemed to rely on both verifying the correct probe as well as rejecting the incorrect probe. On the other hand, the data also showed that the requirement of a carry specificall increased the processing times of the unit digits of the summands for both children and adults consistently. Synced with the findings of the current thesis this suggests that basic principles of place-value processing in mental arithmetic may be acquired early on during learning arithmetic may not change qualitatively. Thereby, the importance of successful mastery of place-value principles in early stages of numerical development (cf. discussion and Figure 1 above) is again emphasized.

However, as already been reported by Miura and colleagues (1989; 1994) there are considerable differences in the accuracy of place-value processing between children but also between different countries and languages. Usually, a superiority of Asian over Western children is observed when it comes to explicitly producing tens and units of a given number. The authors attributed their finding to the fact that the number word systems of most East

Asian languages (e.g., Japanese, Korean) are entirely regular and transparent in their coding of units, tens, hundreds, etc. For instance, 86 is spoken as *hachi-juu-roku* (literally translated: eight-ten-six). Taking into account these differences it remains to be investigated whether the observed association between early place-value understanding and further numerical / arithmetical development represents the general presence and importance of the place-value structuring principle in the world wide dominant Arabic number system. On the other hand, the influence of place-value understanding may be moderated by language specificities such as the inversion property in number words in a way similar to the influence of place-value structuring on the development of the spatial representation of number magnitude (cf. Study 3; Moeller et al., 2009b).

In summary, the results of the current thesis addressing influences of the place-value structure of the Arabic number system on numerical development are straight forward. Extending previous results suggesting place-value information to determine performance in a number line estimation task (cf. Moeller et al., 2009b) it was observed that this influence generalises to language with a number word system without inversion. Nevertheless, the data also indicated that the correspondence between the order of tens and units in a language's number words and symbolical notation serves as a moderating factor of the influence of place-value structuring on the development of the spatial representation of number magnitude. However, apart from this cross-culturaal observation Study 4 of the current thesis also showed that that the influence of the place-value understanding is not limited to basic numerical tasks tapping only basic numerical representations. Instead, longitudinal evidence was reported that indicates a reliable influence of early place-value understanding on later numerical /arithmetical development, in particular, when also requiring processes of place-value integration as in solving carry addition problems. Thus, it can be concluded that on a cross-cultural perspective the place-value ordering principle of the Arabic number system exhibits a strong influence on numerical development in children (cf. Study 3) and that, regarding individual differences, sufficient place-value understanding in early years is a building block for later arithmetical competencies (cf. Study 4).

CONCLUSIONS

All in all, a recapitulated view on the results provided by this thesis reveals a number of important new insights in the influence of the place-value structure of the Arabic number system on numerical cognition. (i) Empirical evidence on decomposed processing of multi-digit numbers obtained from different numerical tasks administered to different populations (i.e., German- and Italian-speaking children, healthy adults, and neglect patients) and assessed using different experimental methodologies (i.e., reaction time, eye-tracking , and fMRI paradigms) further corroborates the notion that the place-value structure of the Arabic number system is retained whenever multi-digit numbers are processed. (ii) Consolidated evaluation of the results of the individual studies provides converging evidence that place-value constraints exhibit a task invariant influence on human numerical cognition and its development. In particular, evidence for the general validity of place-value influences originates from studies revealing (i) that the mental number line is not analogue holistic in nature but incorporates place-value information as well (cf. Studies 1, 5, and 6), (ii) that decomposed processing of two-digit numbers could be observed for new basic numerical as well as basic arithmetical tasks [i.e., number line estimation (cf. Study 3) and addition (cf. Study 2), respectively], (iii) that decomposed processing is reliably influenced by a language's number word system but seems to develop culture invariant (cf. Study 4), (iv) that (computational) models realizing and/or reflecting decomposed processing of tens and units outperformed models of holistic processing when compared directly (cf. Studies 3 and 7), and (v) that on a neurofunctional level specific neural correlates of processing place-value information can be identified. Hence, the results of this thesis can be recapitulated as indicating that decomposed processing of multi-digit numbers complying with the place-value structure of the Arabic number system is a comprehensive, culture invariant but linguistically influenced and developmentally relevant characteristic of human numerical cognition.

REFERENCES

Banks, W. P. (1977). Encoding and processing of symbolic information in comparative judgements. *The Psychology of Learning and Motivation, 11,* 101-159.

Booth, J. L., & Siegler, R. S. (2008). Numerical magnitude representations influence arithmetic learning. *Child Development, 79,* 1016-1031.

Chrisomalis, S. (2004). A cognitive typology for numerical notation. *Cambridge Archaeological Journal, 14,* 37-52.

Cohen, L., & Dehaene, S. (2000). Calculating without reading: Unsuspected residual abilities in pure alexia. *Cognitive Neuropsychology, 17,* 563-583.

Cohen-Kadosh, R., & Walsh, V. (2009). Numerical representation in the parietal lobes: Abstract of not abstract? *Behavioral and Brain Sciences, 32,* 313-328.

Cohen-Kadosh, R., Cohen-Kadosh, K., Kaas, A., Henik, A., & Goebel, R. (2007). Notation-dependent and –independent representations of numbers in the parietal lobes. *Neuron, 53,* 307-314.

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

Dehaene, S., & Cohen, L. (1997). Cerebral Pathways for Calculation: Double Dissociation between Rote Verbal Knowledge and Quantitative Knowledge of Arithmetic. *Cortex, 33,* 219-250.

Dehaene, S., Dupoux, E., & Mehler, J., (1990). Is numerical comparison digital? Analogical and symbolic effects in two-digit number comparison. *Journal of Experimental Psychology: Human Perception and Performance, 16*, 626-641.

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

Delazer, M., Karner, E., Zamarian, L., Donnemiller, E., & Benke, T., 2006. Number processing in posterior cortical atrophy-a neuropsycholgical case study. *Neuropsychologia, 44,* 36-51.

Domahs, F., Nuerk, H.-C., & Delazer, M. (2006). What makes multiplication facts difficult problem Size of neighbourhood consistency? *Experimental Psychology, 53,* 275-282.

Domahs, F., Domahs, U., Schlesewsky, M., Ratinckx, E., Verguts, T., Willmes, K., & Nuerk, H.-C. (2007). Neughborhood consictency in mental arithmetic: Behavioral and ERP evidence. *Behavioral and Brain Functions, 3:* 66.

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

Feigenson, L., Dehaene, S., & Spelke, E. S. (2004). Core systems of number. *Trends in Cognitive Sciences, 8,* 307-314.

Ganor-Stern, D., Pinhas, M., Tzelgov, J. (2008). Comparing two-digit numbers: the importance of being presented together. *Quarterly Journal of Experimental Psychology, 62,* 444-452.

Göbel, S. M., Calabria, M., Farnè, A., & Rossetti, Y. (2006). Parietal rTMS distorts the mental number line: Simulating 'spatial' neglect in healthy subjects. *Neuropsychologia, 44,* 860-868.

Halberda, J., Mazzocco, M. & Feigenson, L. (2008). Individual differences in nonverbal number acuity predict maths achievement. *Nature, 455,* 665-668.

Holloway, I., & 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.

Imbo, I., Vandierendonck, A., & De Rammelaere, S. (2007). The role of working memory in the carry operation of mental arithmetic: Number and value of the carry. *The Quarterly Journal of Experimental Psychology, 60,* 708-731.

Karmiloff-Smith, Annette (1996). *Beyond Modularity: A Developmental Perspective on Cognitive Science*. Cambridge, MA: MIT Press.

Karmiloff-Smith, A. (1998). Development itself is the key to understanding developmental disorders. *Trends in Cognitive Sciences, 2,* 389–398.

Klein, E., Moeller, K., Dressel, K., Domahs, F., Wood, G., Willmes, K., & Nuerk, H.-C. (under revision). To carry or not to carry – is this the question? Disentangling the carry effect in multi-digit addition.

Kong, J., Wang, C., Kwong, K., Vangel, M., Chuac, E., & Gollub, R. (2005). The neural substrate of arithmetic operations and procedure complexity. *Cognitive Brain Research, 22*, 397-405.

Loftus, A. M., Nicholls, M. E. R., Mattingley, J. B. & Bradshaw, J. L. (2008). Numerical processing overcomes left neglect. *Neuroreport, 19,* 835-838.

Macizo, P., & Herrera, A. (in press). Two-digit number comparison: Decade-unit und unit-decade produce the same compatibility effect with number words. *Canadian Journal of Experimental Psychology.*

Macizo, P., Herrera, A., Paolieri, D., & Román, P. (in press). Is there cross-language modulation when bilinguals process number words? *Applied Psycholinguistics.*

Miura, I. T., & Okamoto, Y. (1989). Comparisons of American and Japanese first graders' cognitive representation of number and understanding of place-value. *Journal of Educational Psychology, 81*, 109-113.

Miura, I. T., Okamoto, Y., Kim, C. C., Chang, C.-M., Steere, M., & Fayol, M. (1994). Comparisons of children's cognitive representation of number: China, France, Japan, Korea, Sweden and the United States. *International Journal of Behavioral Development, 17*, 401-411.

Moeller, K. (2006). *Factors affecting behavioural performance and eye movements in the Number Bisection Task.* Unpublished diploma thesis: Technical University Aachen, Aachen, Germany.

Moeller, K., Fischer, M. H., Nuerk, H.-C., & Willmes, K. (2009a). Sequential or parallel decomposed processing of two-digit numbers? Evidence from eye-tracking. *Quarterly Journal of Experimental Psychology, 62,* 323-334.

Moeller, K., Fischer, M. H., Nuerk, H.-C., & Willmes, K. (2009c). Eye fixation behaviour in the number bisection task: Evidence for temporal specificity. *Acta Psychologica, 131,* 209–220.

Moeller, K., Klein, E., & Nuerk, H.-C. (under revision). On the cognitive instantiation of the carry effect in addition – Evidence from eye-tracking.

Moeller, K., Pixner, S., Kaufmann, L., & Nuerk, H.-C. (2009b). Children's early mental number line: Logarithmic or decomposed linear? *Journal of Experimental Child Psychology, 103,* 503-515.

Nieder A., & Dehaene, S. (2009). Representation of number in the brain. *Annual Review in Neuroscience, 32,* 185-208.

Nuerk, H.-C., Weger, U., & Willmes, K. (2001). Decade breaks in the mental number line? Putting the tens and units back in different bins. *Cognition, 82,* B25-33.

Nuerk, H.-C., Geppert, B.E., van Herten, M., & Willmes, K. (2002). On the impact of different number representations in the number bisection task. *Cortex, 38*, 691-715.

Nuerk, H.-C., & Willmes, K. (2005). On the magnitude representations of two-digit numbers. *Psychology Science, 47,* 52-72.

Nuerk, H.-C., Weger, U., & Willmes, K. (2005). Language effects in magnitude comparison: Small, but not irrelevant. *Brain and Language, 92*, 262-277.

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,* 293-305.

Pixner, S. (2009). *How do children process two-digit numbers? Influences of verbal, visual-Arabic and spatial representations on two-digit number processing in children.* Unpublished doctoral thesis, Paris Lodron University, Salzburg, Austria.

Pixner, S., Moeller, K., Zuber, J., & Nuerk, H.-C. (2009). Decomposed but parallel processing of two-digit numbers in 1st graders. *The Open Psychology Journal, 2*, 40-48.

Ratinckx, E., Nuerk, H.-C., van Dijk, J.-P., & Willmes, K. (2006). Effects of interhemispheric communication on two-digit number processing. *Cortex, 42*, 1128-1137.

Rayner, K., & Pollatsek, A. (1989). *The Psychology of Reading.* Englewood Cliffs, NJ: Prentice Hall.

Rubinsten, O., & Henik, A. (2009). Developmental Dyscalculia: heterogeneity might not mean different mechanisms. *Trends in Cognitive Sciences, 13,* 92-99.

Verguts, T., & De Moor, W. (2005). Two-digit comparison decomposed, holistic or hybrid? *Experimental Psychology, 52*, 195-200.

von Aster, M., & Shalev, R. (2007). Number development and developmental dyscalculia. *Developmental Medicine & Child Neurology, 49*, 868-873.

Wood, G., Mahr, M., & Nuerk, H.-C. (2005). Deconstructing and reconstructing the base-10 structure of Arabic numbers. *Psychology Science, 47*, 84-95.

Wood, G., Nuerk, H.-C., & Willmes, K. (2006). Neural representations of two-digit numbers: A parametric fMRI study. *NeuroImage, 46,* 358-367.

Xu, F., Spelke, E., & Goddard, S. (2005). Number sense inhuman infants. *Developmental Science, 8,* 88-101.

Zhang, J., & Wang, H. (2005). The effect of external representation on numeric tasks. *The Quarterly Journal of Experimental Psychology, 58,* 817-838.

Zhou, X., Chen, C., Chen, L., & Dong, Q. (2008). Holistic or compositional representation of two-digit numbers? Evidence from the distance, magnitude, and SNARC effects in a number-matching task. *Cognition, 106,* 1525-1536.

Zorzi, M., Priftis, K., & Umiltà, C. (2002). Neglect disrupts the mental number line. *Nature, 417,* 138-139.

Zorzi, M., Priftis, K., Meneghello, F., Marenzi, R., & Umiltà, C. (2006). The spatial representation of numerical and non-numerical sequences: evidence from neglect. *Neuropsychologia, 44*, 1061-1067.

Zuber, J., Pixner, S., Moeller, K., & Nuerk, H.-C. (2009). On the language-specificity of basic number processing: Transcoding in a language with inversion and its relation to working memory capacity. *Journal of Experimental Child Psychology, 102,* 60-77.

Erklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten oder nicht veröffentlichten Schriften entnommen wurden, sind als solche kenntlich gemacht. Die Arbeit ist in gleicher oder ähnlicher Form oder auszugsweise in einer anderen Prüfung noch nicht vorgelegt worden; auch wurde mit dieser Arbeit oder einer anderen Dissertation noch kein Promotionsversuch unternommen.

Korbinian Möller

Tübingen, den 07.04.2010