

# frontiers

## Frontiers Copyright Statement

© Copyright 2007-2017 Frontiers Media SA. All rights reserved.

All content included on this site, such as text, graphics, logos, button icons, images, video/audio clips, downloads, data compilations and software, is the property of or is licensed to Frontiers Media SA ("Frontiers") or its licensees and/or subcontractors. The copyright in the text of individual articles is the property of their respective authors, subject to a license granted to Frontiers.

The compilation of articles constituting this e-book, wherever published, as well as the compilation of all other content on this site, is the exclusive property of Frontiers. For the conditions for downloading and copying of e-books from Frontiers' website, please see the Terms for Website Use. If purchasing Frontiers e-books from other websites or sources, the conditions of the website concerned apply.

Images and graphics not forming part of user-contributed materials may not be downloaded or copied without permission.

Individual articles may be downloaded and reproduced in accordance with the principles of the CC-BY licence subject to any copyright or other notices. They may not be re-sold as an e-book.

As author or other contributor you grant a CC-BY licence to others to reproduce your articles, including any graphics and third-party materials supplied by you, in accordance with the Conditions for Website Use and subject to any copyright notices which you include in connection with your articles and materials.

All copyright, and all rights therein, are protected by national and international copyright laws.

The above represents a summary only. For the full conditions see the Conditions for Authors and the Conditions for Website Use.

ISSN 1664-8714

ISBN 978-2-88945-200-2

DOI 10.3389/978-2-88945-200-2

## About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

## Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

## Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view.

By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

## What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: [researchtopics@frontiersin.org](mailto:researchtopics@frontiersin.org)

# LINGUISTIC INFLUENCES ON MATHEMATICAL COGNITION

Topic Editors:

**Ann Dowker**, University of Oxford, UK

**Hans-Christoph Nuerk**, University of Tuebingen & LEAD Graduate School and Research Network Tuebingen, Germany

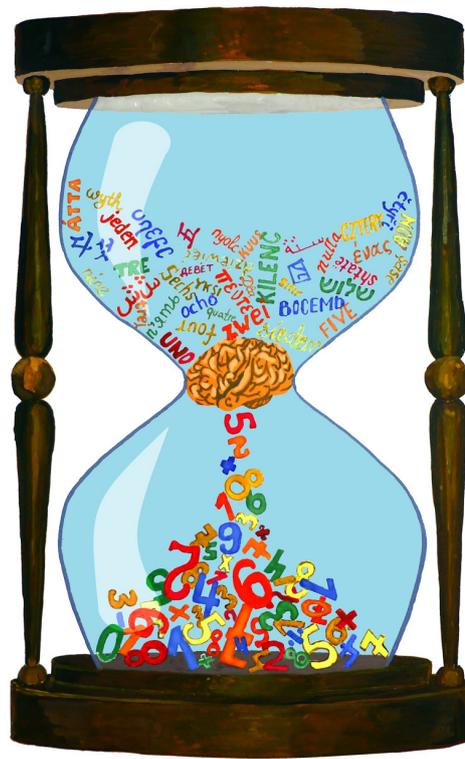


Image by Alexander Artemenko

For many years, an abstract, amodal semantic magnitude representation, largely independent of verbal linguistic representations, has been viewed as the core numerical or mathematical representation. This assumption has been substantially challenged in recent years. Linguistic properties affect not only verbal representations of numbers, but also numerical magnitude representation, spatial magnitude representations, calculation, parity representation, place-value

representation and even early number acquisition. Thus, we postulate that numerical and arithmetic processing are not fully independent of linguistic processing. This is not to say, that in patients, magnitude processing cannot function independently of linguistic processing we just suppose, these functions are connected in the functioning brain. So far, much research about linguistic influences on numerical cognition has simply demonstrated that language influences number without investigating the level at which a particular language influence operates. After an overview, we present new findings on language influences on seven language levels:

- Conceptual: Conceptual properties of language
- Syntactic: The grammatical structure of languages beyond the word level influences
- Semantic: The semantic meaning or existence of words
- Lexical: The lexical composition of words, in particular number words
- Visuo-spatial-orthographic: Orthographic properties, such as the writing/reading direction of a language.
- Phonological: Phonological/phonetic properties of languages
- Other language-related skills: Verbal working memory and other cognitive skills related to language representations

We hope that this book provides a new and structured overview on the exciting influences of linguistic processing on numerical cognition at almost all levels of language processing.

**Citation:** Dowker, A., Nuerk, H-C., eds. (2017). Linguistic Influences on Mathematical Cognition. Lausanne: Frontiers Media. doi: 10.3389/978-2-88945-200-2

# Table of Contents

## **1. Introduction, Structure and Overview**

### **06 Editorial: Linguistic Influences on Mathematics**

Ann Dowker and Hans-Christoph Nuerk

## **2. Conceptual influences**

### **10 Arbitrary numbers counter fair decisions: trails of markedness in card distribution**

Philipp A. Schroeder and Roland Pfister

## **3. Syntactic influences**

### **18 On the relation between grammatical number and cardinal numbers in development**

Barbara W. Sarnecka

## **4. Semantic influences**

### **22 Word problems: a review of linguistic and numerical factors contributing to their difficulty**

Gabriella Daroczy, Magdalena Wolska, Walt Detmar Meurers and Hans-Christoph Nuerk

## **5. Lexical influences**

### **5.1. Lexical transparency: The case of inversion**

#### **35 Intransparent German number words complicate transcoding – a translingual comparison with Japanese**

Korbinian Moeller, Julia Zuber, Naoko Olsen, Hans-Christoph Nuerk and Klaus Willmes

#### **45 The developmental onset of symbolic approximation: beyond nonsymbolic representations, the language of numbers matters**

Iro Xenidou-Dervou, Camilla Gilmore, Menno van der Schoot and Ernest C. D. M. van Lieshout

#### **58 On the limits of language influences on numerical cognition – no inversion effects in three-digit number magnitude processing in adults**

Julia Bahnmueller, Korbinian Moeller, Anne Mann and Hans-Christoph Nuerk

#### **68 The relation between language and arithmetic in bilinguals: insights from different stages of language acquisition**

Amandine Van Rinsveld, Martin Brunner, Karin Landerl, Christine Schiltz and Sonja Ugen

#### **83 Number word structure in first and second language influences arithmetic skills**

Anat Prior, Michal Katz, Islam Mahajna and Orly Rubinsten

### **5.2. Lexical transparency: The case of power transparency**

#### **93 Does the transparency of the counting system affect children's numerical abilities?**

Ann Dowker and Manon Roberts

- 100** *Linguistic influence on mathematical development is specific rather than pervasive: revisiting the Chinese Number Advantage in Chinese and English children*  
Winifred Mark and Ann Dowker
- 109** *Corrigendum: Linguistic influence on mathematical development is specific rather than pervasive: revisiting the Chinese Number Advantage in Chinese and English children*  
Winifred Mark and Ann Dowker

## **6. Visuo-spatial-orthographic influences**

### **6.1. Spatial direction of script**

- 110** *How space-number associations may be created in preliterate children: six distinct mechanisms*  
Hans-Christoph Nuerk, Katarzyna Patro, Ulrike Cress, Ulrike Schild, Claudia K. Friedrich and Silke M. Göbel
- 116** *Up or down? Reading direction influences vertical counting direction in the horizontal plane – a cross-cultural comparison*  
Silke M. Göbel
- 128** *Two steps to space for numbers*  
Martin H. Fischer and Samuel Shaki

### **6.2. Spatial complexity of script**

- 131** *Spatial complexity of character-based writing systems and arithmetic in primary school: a longitudinal study*  
Maja Rodic, Tatiana Tikhomirova, Tatiana Kolienco, Sergey Malykh, Olga Bogdanova, Dina Y. Zueva, Elena I. Gynku, Sirui Wan, Xinlin Zhou and Yulia Kovas

## **7. Phonological or auditory influences**

- 142** *Mathematics and reading difficulty subtypes: minor phonological influences on mathematics for 5–7-years-old*  
Julie A. Jordan, Judith Wylie and Gerry Mulhern
- 154** *Number processing and arithmetic skills in children with cochlear implants*  
Silvia Pixner, Martin Leyrer and Korbinian Moeller

## **8. Other language-related influences: Verbal working memory**

- 164** *Contribution of working memory in multiplication fact network in children may shift from verbal to visuo-spatial: a longitudinal investigation*  
Mojtaba Soltanlou, Silvia Pixner and Hans-Christoph Nuerk



# Editorial: Linguistic Influences on Mathematics

Ann Dowker<sup>1\*</sup> and Hans-Christoph Nuerk<sup>2,3,4</sup>

<sup>1</sup> Experimental Psychology, University of Oxford, Oxford, UK, <sup>2</sup> Department of Psychology, University of Tuebingen, Tuebingen, Germany, <sup>3</sup> Knowledge Media Research Center, University of Tuebingen, Tuebingen, Germany, <sup>4</sup> LEAD Graduate School and Research Network, Tuebingen, Germany

**Keywords:** language, numerical cognition, psychology of arithmetic, verbal counting systems, cross-linguistic research

## The Editorial on the Research Topic

### Linguistic Influences on Mathematics

For many years, an abstract, amodal semantic magnitude representation, largely independent of verbal linguistic representations, has been viewed as the core numerical or mathematical representation (Dehaene and Cohen, 1995). This assumption has been substantially challenged in recent years (e.g., Miura and Okamoto, 2003; Nuerk et al., 2004, 2005; Dowker et al., 2008; Colomé et al., 2010; Helmreich et al., 2011; Krinzinger et al., 2011; Pixner et al., 2011a,b; Göbel et al., 2014; Imbo et al.; Klein et al.). Linguistic properties affect not only verbal representations of numbers (Seron and Fayol, 1994; Zuber et al., 2009; Pixner et al., 2011a), but also numerical magnitude representation (Nuerk et al., 2005; Pixner et al., 2011b), spatial magnitude representations (Shaki et al., 2009; Helmreich et al., 2011), calculation (Colomé et al., 2010; Krinzinger et al., 2011; Göbel et al., 2014), parity representation (Iversen et al., 2004, 2006; Nuerk et al., 2004), place-value representation (Miura and Okamoto, 2003; for a review, see Nuerk et al.) and even early number acquisition (Sarnecka, this issue). Thus, we postulate that numerical and arithmetic processing are not fully independent of linguistic processing. This is not to say, that in patients, magnitude processing cannot function independently of linguistic processing (e.g., Dehaene and Cohen, 1997), we just suppose, these functions are connected in the functioning brain. So far, much research about linguistic influences on numerical cognition has simply demonstrated that language influences number without investigating the level at which a particular language influence operates. Here we want to distinguish several linguistic levels at which numerical processing may be influenced, according to which we group the articles in our special issue:

## OPEN ACCESS

**Edited and reviewed by:**  
Jessica S. Horst,  
University of Sussex, UK

**\*Correspondence:**  
Ann Dowker  
ann.dowker@psy.ox.ac.uk

**Specialty section:**  
This article was submitted to  
Developmental Psychology,  
a section of the journal  
Frontiers in Psychology

**Received:** 09 May 2016  
**Accepted:** 24 June 2016  
**Published:** 12 July 2016

**Citation:**  
Dowker A and Nuerk H-C (2016)  
Editorial: Linguistic Influences on  
Mathematics. *Front. Psychol.* 7:1035.  
doi: 10.3389/fpsyg.2016.01035

- **Conceptual:** Conceptual properties of language
- **Syntactic:** The grammatical structure of languages beyond the word level influences
- **Semantic:** The semantic meaning or existence of words
- **Lexical:** The lexical composition of words, in particular number words
- **Visuo-spatial-orthographic:** Orthographic properties, such as the writing/reading direction of a language.
- **Phonological:** Phonological/phonetic properties of languages
- **Other language-related skills:** Verbal working memory and other cognitive skills related to language representations

## CONCEPTUAL INFLUENCES

Beyond single phonemes, graphemes, words and sentences, linguistic structures are also shaped by linguistic concepts. The linguistic markedness concept suggests that for (almost) each adjective pair, a ground (unmarked) form and a derived (marked) form exist (e.g., efficient and inefficient; marked by “in”).

We consider the markedness concept “conceptual” (see Nuerk et al., 2004). However, many language models do not consider a conceptual level as such and often the lexical or semantic level is the highest level. Levelt et al. (1999), however, proposed a conceptual level in the language production model. It is the highest level in this model and is assumed to be involved in the conceptual preparation of lexical concepts. In Nuerk et al. (2004, p.859), we suggested that linguistic markedness could operate at just such a conceptual level and that other verbal influences like phonological ones will operate at a different (lower) level, e.g., the phonological encoding in the mental lexicon.

Numbers possess several attributes, which can be distinguished into unmarked ground form (large, even, divisible) and marked form (small, odd, indivisible; Hines, 1990). As regards spatial organization “right” is unmarked and “left” is marked (Nuerk et al., 2004). Usually responses are faster, when markedness of stimuli and responses are congruent (e.g., left-odd, right-even). Schroeder and Pfister (this issue) investigated SNARC and MARC effects on card distribution to fellow card players. They observed markedness effects in that magnitude and parity influence card distribution. However, in this natural setting, the markedness effect is inverted to a normal parity judgment task, extending earlier findings in deaf signers (Iversen et al., 2004), and left-handers (Huber et al., 2015). This implies that not only bodily, but also task-specific constraints need to be taken into account, when linguistic effects on mathematical cognition on the construct level are examined.

## SYNTACTIC INFLUENCES

Number processing in real life situations occurs in natural language and is described by grammatical number. (i.e., singular for 1 and plural for numbers 2 and greater in English). Languages differ substantially in their use of grammatical number (see Overmann, 2015) analysis of 905 languages): For instance, 7% of these languages lacked grammatical number altogether despite having lexical numbers. Influences of grammatical numbers on numerical cognition have been shown in two effects. First, Roettger and Domahs (2014) observed a grammatical SNARC effect: singular inflected words elicited faster responses on the left hand side and plural inflected words on the right. Second, as beautifully outlined by Sarnecka’s (this issue) review, the sheer existence of certain grammatical number enhances development of number concepts in children. In languages without differentiation between singular and plural, the development of number understanding in children is later. Moreover, grammatical distinction between singular, dual (a grammatical form for “two”) and plural present in several languages further enhances, yet partially hinders number development in children. In some cases, the syntactic structure of a language both influences development of numerical understanding and spatial mappings of numbers.

## SEMANTIC INFLUENCES

Word meanings also influence numerical or arithmetic processing. Daroczy et al. reviewed text problems and

found that numerical properties and semantic properties are often interacting. For example, the consistency effect suggests that text problems are easier, when the required operation is consistently associated with the semantics of the words. For instance, addition is more associated with “more,” “buy,” “get,” etc., while subtraction is more associated with “less,” “sell,” “give,” etc. When text problems are presented in a way that makes such associations misleading, children and adults perform less well. This highlights an interrelation between word meaning and preferred arithmetic operations.

## LEXICAL INFLUENCES

Most of the papers in our special issue as well as in the literature are concerned with lexical influences, in particular number words. In general, a transparent number word structure seems to help numerical performance even for problems not involving number words (Nuerk et al., 2015). Two types of lexical influences are discussed in our special issue. The first involves the inversion property. Some languages like Arabic, Dutch and German invert the order of tens and units (“one-and-twenty” for 21), which creates problems in several tasks. Moeller et al. (this issue) compared transcoding (writing numbers to dictation) skills in Japanese and German. The Japanese children did much better. In particular, Japanese children make far fewer inversion errors; but also fewer errors in general. Xenidou-Dervou et al. (this issue) show that the inversion property does not affect all numerical and arithmetic skills. Dutch children (with inversion) lag behind English children in symbolic but not non-symbolic arithmetic. A working memory overload in Dutch was found in non-symbolic, but not symbolic magnitude. However, as Bahnmüller et al. (this issue) show, inversion effects do not even affect all aspects of symbolic number processing. While children’s and adults’ two-digit Arabic number comparison is influenced by inversion properties of a language, adults’ three-digit Arabic number comparison is not. Moreover, van Rinsveld et al. (this issue) found that inversion affected complex but not simple symbolic arithmetic in German-French bilingual secondary pupils. Finally, Prior et al. (this issue) gave Hebrew-Arabic bilinguals oral arithmetic problems, because Arabic but not Hebrew number words possess the inversion property. Participants solved arithmetic problems best when the language structure corresponded to the arithmetic problem. This implies that—contrary to earlier claims—L1 does not completely dominate arithmetic processing, but that both L1 and L2 shape numerical and arithmetic.

The second line of research at the lexical level is power transparency. Unlike most European languages, most Asian languages are extremely transparent with respect to the power of a given number (e.g., “ten-two” for 12). From 11 on, children and adults can derive the power of each number directly from the number word. It has been argued that this transparency may be responsible for Asians’ better skills at counting, representing 2-digit numbers, and general arithmetic (Miller et al., 1995; Miura and Okamoto, 2003). However, such results are confounded by the many other educational and

cultural differences between countries. One way of obtaining more specific evidence of language effects is to compare children studying in different languages in the same country and educational system. For instance, the Welsh counting system, unlike the English system, is transparent. Dowker et al. (2008) found that children in Welsh-medium primary schools did not do better in arithmetic overall, but showed specific advantages in reading and comparing two-digit numbers. Extending those results Dowker and Roberts observed that Welsh-medium children give more precise and consistent representations of 2-digit numbers on empty number line tasks. Mark and Dowker studied children in Chinese and English medium primary schools in Hong Kong. The Chinese medium children were better at some tasks but not others: e.g., they were better at counting backwards but not forwards; and were not better at number comparison. Thus, we can conclude that lexical influences do affect arithmetic, but not as pervasively as sometimes assumed.

## VISUO-SPATIAL-ORTHOGRAPHIC INFLUENCES

Visual-spatial-orthographic influences mostly involve the reading/writing direction of a given script or its complexity. Usually, space-number relations are associated with the dominant reading/writing direction (for a review see Fischer and Shaki, 2014). However, reading/writing direction already influences spatial-numerical directionality, before children can read or write (Patro and Haman, 2012; Nuerk et al., 2015). Most studies so far have investigated visuo-spatial-orthographic influences on the horizontal left/right dimension. Göbel (this issue) showed that cultural influences on number-space-relations also include the vertical dimension. Fischer and Shaki (this issue) proposed two steps in the shaping of directional space-number representations in adults: “*the spatial dimension selected for mapping of numbers reflects the stimulus and response features of the current task*” and “*the orientation of the SNA is influenced by spatial experience.*”

Relatedly, Rodic et al. examined whether learning spatially complex scripts (e.g., Chinese) is related to mathematical performance. They found no evidence that exposure to a spatially complex script improves mathematics.

We conclude that visuo-spatial orthographic skills seem to shape the direction of space-number relations, but not arithmetic skills themselves.

## REFERENCES

- Ashcraft, M. H., and Stazyk, E. H. (1981). Mental addition: a test of three verification models. *Mem. Cognit.* 9, 185–196. doi: 10.3758/BF03202334
- Colomé, Á., Laka, I., and Sebastián-Gallés, N. (2010). Language effects in addition: how you say it counts. *Q. J. Exp. Psychol.* 63, 965–983. doi: 10.1080/17470210903134377
- Dehaene, S., and Cohen, L. (1995). Towards an anatomical and functional model of number processing. *Math. Cogn.* 1, 83–120.

## PHONOLOGICAL INFLUENCES

Jordan et al. examined phonological skills in children with difficulties in reading, mathematics or both and found minor influences on phonology on mathematics. Pixner et al. (this issue) examined children with cochlear implants (CI), who usually have phonological (and also other) language deficits. They found general deficits in such children in multiplication, subtraction and number line estimation, but specific deficits in (verbally mediated) place-value manipulation. We conclude that phonological skills are not related to mathematical functioning *per-se*, but to verbal representations/manipulations of number.

## OTHER LANGUAGE-RELATED SKILLS: VERBAL WORKING MEMORY AND OTHER COGNITIVE SKILLS

Verbal working memory is associated with complex arithmetic since Ashcraft and Stazyk (1981) seminal paper. Soltanlou et al. (this issue) investigated whether verbal or spatial working memory influences multiplication skill most strongly. They observed an age-related shift from verbal WM to spatial WM influences over time. Thus, working memory data from adults or one children age-group are not representative for its influence in different developmental stages.

## SUMMARY

Linguistic influences on number processing are ubiquitous. They occur at conceptual, semantic, syntactic, lexical, visuo-spatial-orthographic, phonological, and other levels. Research should now address more precisely which language characteristics at which level influence particular numerical tasks at particular ages.

## AUTHOR CONTRIBUTIONS

All authors listed, have made substantial, direct and intellectual contribution to the work, and approved it for publication.

## ACKNOWLEDGMENTS

HN work was supported by funding of the German Research Foundation (DFG NU 265/3-1) on “Linguistic Influences on Numerical Cognition: A cross-cultural investigation using natural specificities of Polish and German languages.”

- Dehaene, S., and Cohen, L. (1997). Cerebral pathways for calculation: double dissociation between rote verbal and quantitative knowledge of arithmetic. *Cortex* 33, 219–250. doi: 10.1016/S0010-9452(08)70002-9
- Dowker, A., Bala, S., and Lloyd, D. (2008). Linguistic influences on mathematical development: how important is the transparency of the counting system? *Philos. Psychol.* 21, 523–538. doi: 10.1080/09515080802285511
- Fischer, M. H., and Shaki, S. (2014). Spatial associations in numerical cognition—From single digits to arithmetic. *Q. J. Exp. Psychol.* 67, 1461–1483. doi: 10.1080/17470218.2014.927515

- Göbel, S. M., Moeller, K., Pixner, S., Kaufmann, L., and Nuerk, H.-C. (2014). Language affects symbolic arithmetic in children: the case of number word inversion. *J. Exp. Child Psychol.* 119, 17–25. doi: 10.1016/j.jecp.2013.10.001
- Helmreich, I., Zuber, J., Pixner, S., Kaufmann, L., Nuerk, H.-C., and Moeller, K. (2011). Language effects on children's non-verbal number line estimations. *J. Cross Cult. Psychol.* 42, 598–613. doi: 10.1177/0022022111406026
- Hines, T. M. (1990). An odd effect: lengthened reaction times for judgments about odd digits. *Mem. Cognit.* 18, 40–46. doi: 10.3758/BF03202644
- Huber, S., Klein, E., Graf, M., Nuerk, H.-C., Moeller, K., and Willmes, K. (2015). Embodied markedness of parity? Examining handedness effects on parity judgments. *Psychol. Res.* 79, 963–977. doi: 10.1007/s00426-014-0626-9
- Iversen, W., Nuerk, H.-C., Jäger, L., and Willmes, K. (2006). The influence of an external symbol system on number parity representation, or what's odd about 6? *Psychon. Bull. Rev.* 13, 730–736. doi: 10.3758/BF03193988
- Iversen, W., Nuerk, H.-C., and Willmes, K. (2004). Do signers think differently? The processing of number parity in deaf participants. *Cortex* 40, 176–178. doi: 10.1016/S0010-9452(08)70940-7
- Krinzinger, H., Gregoire, J., Desoete, A., Kaufmann, L., Nuerk, H.-C., and Willmes, K. (2011). Effects of language, curricula on numerical skills in second grade. *J. Cross Cult. Psychol.* 42, 614–662. doi: 10.1177/0022022111406252
- Levelt, W. J. M., Roelofs, A., and Meyer, A. S. (1999). A theory of lexical access in speech production. *Behav. Brain Sci.* 22, 1–38. doi: 10.1017/s0140525x99001776
- Miller, K. F., Smith, C. M., Zhu, J., and Zhang, H. (1995). Preschool origins of cross-national differences in mathematical competence: the role of number-naming systems. *Psychol. Sci.* 6, 56–60. doi: 10.1111/j.1467-9280.1995.tb00305.x
- Miura, I. T., and Okamoto, Y. (2003). "Language supports for mathematics understanding and performance." in *The Development of Arithmetic Concepts and Skills*, eds A. Baroody and A. Dowker (Mahwah, NJ: Lawrence Erlbaum Associates, Inc.), 229–242.
- Nuerk, H.-C., Iversen, W., and Willmes, K. (2004). Notational modulation of the SNARC and the MARC (linguistic markedness of response codes) effect. *Q. J. Exp. Psychol. Section A* 57, 835–863. doi: 10.1080/02724980343000512
- Nuerk, H.-C., Moeller, K., and Willmes, K. (2015). "Multi-digit number processing – overview, conceptual clarifications, and language influences," in *The Oxford Handbook of Numerical Cognition*, eds R. Cohen Kadosh and A. Dowker (Oxford: Oxford University Press), 106–139.
- Nuerk, H.-C., Weger, U., and Willmes, K. (2005). Language effects in magnitude comparison: small, but not irrelevant. *Brain Lang.* 92, 262–277. doi: 10.1016/j.bandl.2004.06.107
- Overmann, K. A. (2015). Numerosity structures the expression of quantity in lexical numbers and grammatical number. *Curr. Anthropol.* 56, 638–653. doi: 10.1086/683092
- Patro, K., and Haman, M. (2012). The spatial-numerical congruity effect in preschoolers. *J. Exp. Child Psychol.* 111, 534–542. doi: 10.1016/j.jecp.2011.09.006
- Pixner, S., Moeller, K., Heřmanová, V., Nuerk, H.-C., and Kaufmann, L. (2011a). Whorf reloaded: language effects on nonverbal number processing in first grade—A trilingual study. *J. Exp. Child Psychol.* 108, 371–382. doi: 10.1016/j.jecp.2010.09.002
- Pixner, S., Zuber, J., Hermanová, V., Kaufmann, L., Nuerk, H.-C., and Moeller, K. (2011b). One language, two number-word systems and many problems: numerical cognition in the Czech language. *Res. Dev. Disabil.* 32, 2683–2689. doi: 10.1016/j.ridd.2011.06.004
- Roettger, T. B., and Domahs, F. (2014). Grammatical number elicits spatial-numerical association of response codes (SNARC) and linguistic markedness of response codes (MARC) effects as a function of task demands. *Psychology* 57, 835–863. doi: 10.1080/17470218.2014.979843
- Seron, X., and Fayol, M. (1994). Number transcoding in children: a functional analysis. *Br. J. Dev. Psychol.* 12, 281–300. doi: 10.1111/j.2044-835X.1994.tb00635.x
- Shaki, S., Fischer, M. H., and Petrusic, W. M. (2009). Reading habits for both words and numbers contribute to the SNARC effect. *Psychon. Bull. Rev.* 16, 328–331. doi: 10.3758/PBR.16.2.328
- Zuber, J., Pixner, S., Moeller, K., and 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. *J. Exp. Child Psychol.* 102, 60–77. doi: 10.1016/j.jecp.2008.04.003

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Dowker and Nuerk. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Arbitrary numbers counter fair decisions: trails of markedness in card distribution

Philipp A. Schroeder<sup>1\*</sup> and Roland Pfister<sup>2</sup>

<sup>1</sup> Department of Psychiatry and Psychotherapy, Neurophysiology and Interventional Neuropsychiatry, University of Tübingen, Tübingen, Germany

<sup>2</sup> Department of Psychology III, University of Würzburg, Würzburg, Germany

## Edited by:

Hans-Christoph Nuerk, University of Tübingen, Germany  
Ann Dowker, University of Oxford, UK

## Reviewed by:

Stefan Huber, Knowledge Media Research Center, Germany  
Tobias Loetscher, University of South Australia, Australia

## \*Correspondence:

Philipp A. Schroeder, Department of Psychiatry and Psychotherapy, Neurophysiology and Interventional Neuropsychiatry, University of Tübingen, Calwerstrasse 14, D-72076 Tübingen, Germany  
e-mail: philipp.schroeder@uni-tuebingen.de

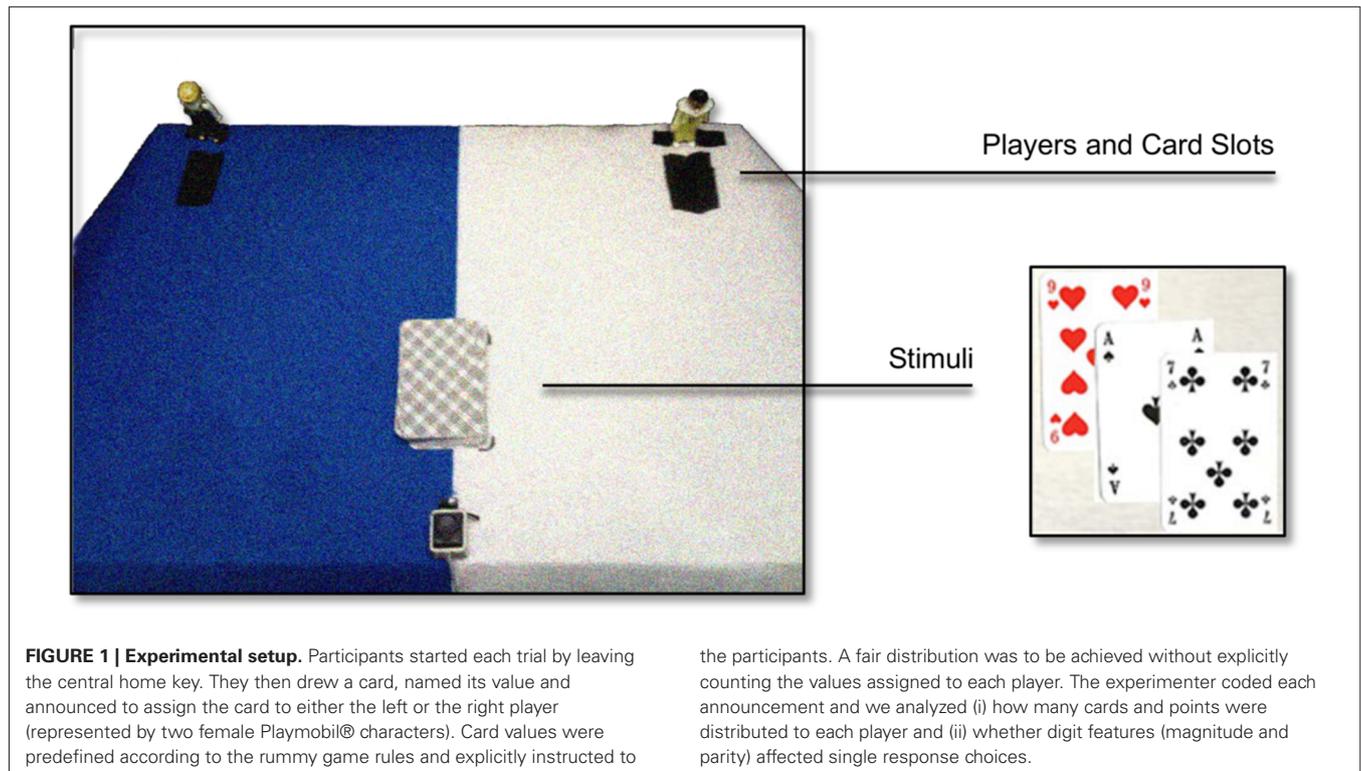
Converging evidence from controlled experiments suggests that the mere processing of a number and its attributes such as value or parity might affect free choice decisions between different actions. For example the spatial numerical associations of response codes (SNARC) effect indicates the magnitude of a digit to be associated with a spatial representation and might therefore affect spatial response choices (i.e., decisions between a “left” and a “right” option). At the same time, other (linguistic) features of a number such as parity are embedded into space and might likewise prime left or right responses through feature words [odd or even, respectively; markedness association of response codes (MARC) effect]. In this experiment we aimed at documenting such influences in a natural setting. We therefore assessed number-space and parity-space association effects by exposing participants to a fair distribution task in a card playing scenario. Participants drew cards, read out loud their number values, and announced their response choice, i.e., dealing it to a left vs. right player, indicated by Playmobil characters. Not only did participants prefer to deal more cards to the right player, the card’s digits also affected response choices and led to a slightly but systematically unfair distribution, supported by a regular SNARC effect and counteracted by a reversed MARC effect. The experiment demonstrates the impact of SNARC- and MARC-like biases in free choice behavior through verbal and visual numerical information processing even in a setting with high external validity.

**Keywords:** embodied cognition, numerical cognition, SNARC effect, MARC effect, and justice for all, linguistic markedness, free choice

## INTRODUCTION

Like nothing else, numbers are regarded as pure and objective. They are the cornerstone of scientific progress in terms of measurements and statistics and they similarly shape global business in various ways—from defining monthly salaries to describing trends at the stock market. But does this objectivity survive when numbers come in contact with human agents? In fact, there seems to be good reason for a positive answer to this question. Numbers obviously allow for rule-based decisions between competing options, and a decision that is based on numbers is readily accepted as fair and impersonal (Porter, 1996). At the same time, however, research on human decision making has documented that numbers can systematically bias an agent’s choice behavior via anchoring and adjustment heuristics (Mussweiler and Englich, 2003; Furnham and Boo, 2011). For instance, when asked to estimate the value of a property, laymen and professionals alike rated the price of a real estate higher when they were told a higher listed price before (Northcraft and Neale, 1987). This anchoring bias was found in numerous contexts and research in this domain has shown that heuristic decisions might even integrate nominally irrelevant anchors like telephone and social insurance numbers (Tversky and Kahneman, 1974).

Such anchoring effects are of course driven by memory processes rather than by the numbers themselves. Still, recent research on numerical and embodied cognition suggests that the mere presence of a number alone might be sufficient to invoke biases in thoughts and actions (Barsalou, 1999; Fischer, 2006, 2012). These biases built on well-documented associations between numerical magnitude and spatial locations that indicate smaller numbers to be associated with left locations and larger numbers to be associated with right locations [spatial numerical associations of response codes (SNARC) effect; Dehaene et al., 1993; Wood et al., 2008]. Most importantly for the present study, such spatial-numerical associations also affect response choices (Tschemtscher et al., 2012; Shaki and Fischer, 2014). That is, when being confronted with smaller numbers, participants showed a preference for choosing a left vs. a right response key (Daar and Pratt, 2008) and, similarly, such small numbers involuntarily prompted left-oriented gaze directions (Ruiz Fernández et al., 2011) and small numbers were produced more likely while turning or gazing to the left (Loetscher et al., 2008, 2010). These automatic biases document that the mere presence of a number is sufficient to bias choices and behavior. Sensory and motor biases induced by the SNARC effect can be considered of high



diagnostic merit for the understanding of grounded, embodied, and situated cognition (Fischer, 2012). Findings pertinent to this point range from culture-dependent finger counting habits that influence magnitude representations (Domahs et al., 2010) to bodily postures (Eerland et al., 2011) or even “unusual bodies” (Keehner and Fischer, 2012) that introduce peculiarities in spatial tasks. Together, these studies indicate that numerical associations reliably alter spatial response choices in deliberately employed highly controlled settings where the agent does not pursue any other goals except for deciding spontaneously for a spatially coded response.

As a first aim, the present study investigated whether the described bias would also occur in a more externally valid setting such as in situations where the agent aims at fairly and objectively distributing value among other people. We operationalized this situation in terms of a card distribution task in which participants were asked to deal cards of a given value to a player to the left or to the right and additionally announce their value-space choice (Figure 1). If spatial-numerical biases do indeed generalize to this situation, participants should deal more cards with higher values to the right player than to the left player.

Of course, these biases do not work in an all or none fashion, but gradually. That is, even though participants prefer choices that are congruent to a number’s spatial association (e.g., a left response to a small number), they also tend to show a fair amount of incongruent choices (e.g., a right response to a small number; Daar and Pratt, 2008). In the natural card playing setting of this study, however, both spatial-numerical associations and markedness of parity and space [markedness association of response codes (MARC) effect; Nuerk et al., 2004] might affect choice

probabilities for each single card, summing up to an overall biased and therefore unfair bias in value distribution. As both, high and even numbers (such as the target card value “8” or “10”) are usually associated with right responses and with more points in the rummy card setting at hand, our main hypothesis was that participants would be biased to deal overall more points to the right than to the left player.

## MATERIALS AND METHODS

### PARTICIPANTS AND APPARATUS

Twenty-five participants (19 females, mean age = 24.3, range: 18–52 years, 3 left-handed)<sup>1</sup> were invited to participate in a 15-min experimental session. They were seated in front of the apparatus displayed in Figure 1. This apparatus mainly consisted of a 60 × 40 cm cardboard box, the surface of which was covered with blue and white paper. Two Playmobil® characters represented the players and were positioned at the rear edge of the card box surface with an inter-player distance of 50 cm. The players were matched for various attributes such as size, age, beauty, and orientation toward the participant. A slot in front of each player allowed the participants to insert a card in a box beneath the surface of the apparatus, restricting visual feedback of the current distribution. A central key was positioned at the front edge to allow for a standardized trial procedure, and the

<sup>1</sup>As pointed out by a reviewer, individual variations of age, handedness or sex might play a role in marked decisions about numbers (see the discussion for an elaboration). However, also following the reviewer’s suggestion, fitting a model on right-handed female participants aged 30 or less did not substantially alter the results and only marginally improved the model fit.

card deck was placed 10 cm from the key onto a predefined mark. One participant decided to distribute cards by color and thereby achieved a totally fair distribution; this participants' data was excluded from the analysis and we refer to the remaining  $N = 24$  participants in the following. The study was conducted in accordance with the Declaration of Helsinki and the guidelines of the ethics committee at the University of Würzburg.

## PROCEDURE

The basic task of the participants was to draw a card and deal it to either the left or the right player. Each participant received three random training cards, then a complete 52 Anglo-American style rummy card pack. We ensured that the card icons were printed in all four corners of each card to avoid systematic influences originating from the specific stimulus set (Figure 1). Card values were defined following standard rummy game rules, that is: number cards (2–10) counted their printed value (i.e., two points for a “2,” three points for a “3,” and so on), royal cards (jack, queen, and king) counted 10 points, and aces counted 11 points. The deck was professionally shuffled prior to the experiment. During the instructions, we emphasized that participants should aim for a fair distribution of values across players by intuition and without using any explicit strategies (such as counting points across the experiment).

To start a trial, participants pressed and released the start button. They then drew the top card from the deck, read out loud the card's face (e.g., “Ace of Spades”), its value (“11”), and announced the side they wanted to distribute it to (always in this order). They then inserted the card into the right or left card slot. The experimenter registered the information and also coded invalid trials (i.e., illegal use of the left hand, reading out the wrong number or value, or naming the card's attributes and the corresponding choice in the wrong order; 4.4% trials in total).

## DATA TREATMENT

For the main analysis, both the number of cards and the resulting scores for each player and participant were computed. Note that although the two measures are confounded, they still allow for distinct evaluation of choice preference and influences of the SNARC or the MARC effect: Even without an overall preference of one player in terms of the number of cards, a difference in scores can arise from a SNARC-like distribution of high-value cards to the right player and low-value cards to the left player. Both measures were controlled for homogeneity and normal distribution and subjected to one-tailed paired  $t$ -tests to assess our main hypothesis of a preference for the right player.

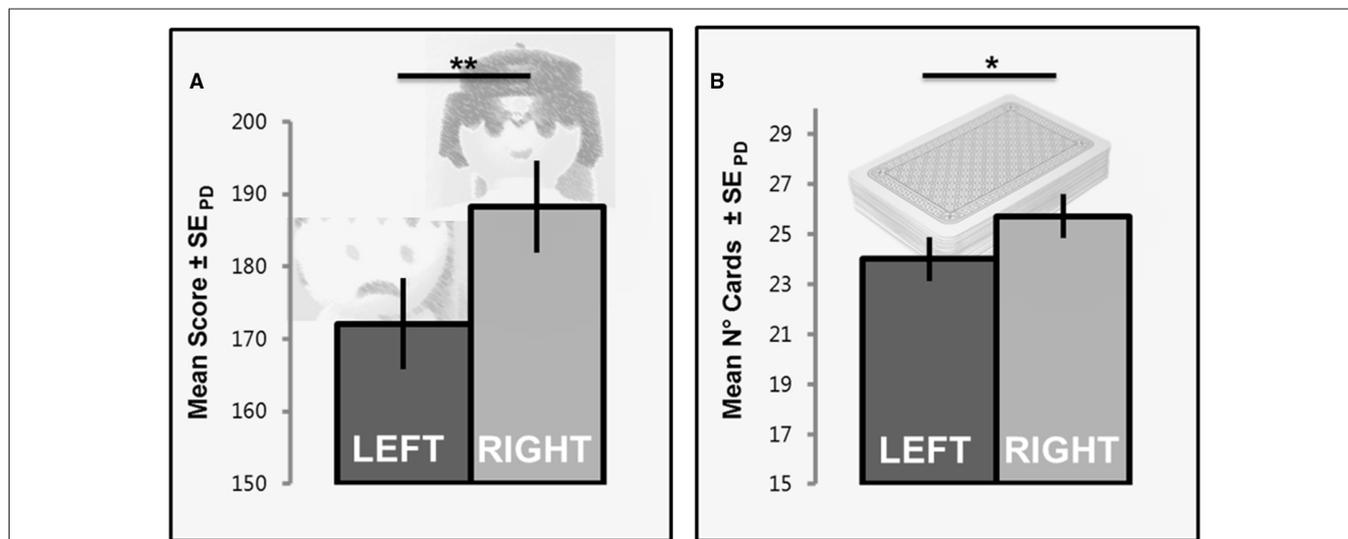
In a second, exploratory analysis, we aimed at dismantling underlying SNARC and MARC influences to the free, binary choice at a trial-wise level. Therefore, we used generalized mixed-effects models to predict the likelihood of a left response from the two first-level fixed factors parity and magnitude.

## RESULTS

### SCORES AND NUMBER OF CARDS

Mean scores and number of cards for each player are depicted in Figure 2. Tests for normal distribution (Kolmogorov Smirnov:  $p > 0.23$ ) and homogeneity of the sample were conducted prior to the analysis and showed the data to be suitable for analyses via parametric tests.

Whereas 188 (SE = 3.33) points on average were assigned to the right player, only 172 (SE = 3.76) points were assigned to the left player, and this difference in scores was significant,  $t(23) = 2.52$ ,  $p = 0.010$ ,  $d = 0.53$  (Figure 2A). A similar effect emerged for the number of cards dealt to the left and right player, respectively,  $t(23) = 1.92$ ,  $p = 0.034$ ,  $d = 0.40$  (Figure 2B), as participants assigned about two cards more ( $\bar{d}_N = 1.71$ , SE = 0.48) to the right player. The effects on points and card numbers were



**FIGURE 2 | Mean scores and standard errors of paired difference (cf. Pfister and Janczyk, 2013) in the card distribution task.** Participants overall preferred the right player which resulted in a significant difference in scores

(A), and a similar effect in overall card numbers (B). Note that both Playmobil players acted earnest without any particular facial expression during the experiment, unlike the displayed emotions in panel (A). \* $p < 0.05$ , \*\* $p < 0.01$ .

**Table 1 | Probabilities of left response choices as a function of target card value.**

Target card value P(left) [%]	Single-digit cards								Royal cards	
	2	3	4	5	6	7	8	9	10	11
	57.3	50.4	48.3	42.0	52.8	40.6	48.5	42.4	49.4	47.2

correlated significantly across participants,  $r = 0.85$ ,  $p < 0.001$ , indicating that the difference in cards accounted for about 71% of the effect on distributed points.

### EXPLORATORY ANALYSIS: SNARC AND MARC EFFECTS

More fine-grained analyses targeted the outcome of individual decisions rather than the overall number of points or cards dealt by each participant (Table 1). More precisely we aimed at analyzing the impact of magnitude and parity on the outcome of a decision (i.e., the likelihood for a card to be dealt to the left or to the right). To this end, we employed generalized linear mixed-effects models to model the binary outcome of the choice. Magnitude and parity were entered as fixed factors into the model (first level predictors), which further included individual subjects as random effects on the second level. The model was fitted in R by using the *glmer* function of the *lme4* package (Bates et al., 2014; binomial family and logit link function). We further restricted the analysis to single-digit values (2–9) due to the actual different pictorial presentation of royal card values and possibly different representational format of values that would imply a two-digit numerical notation (Nuerk and Willmes, 2005; Nuerk et al., 2011).

In a first step, we evaluated each predictor individually (each being coded as centered variable). As suggested by the main analyses above, higher magnitudes were indeed associated with a higher preference for right responses (fixed effect estimate = 0.055/number, SE = 0.032),  $z = 1.70$ ,  $p = 0.045$ ,  $p_{pb} = 0.036$ .<sup>2</sup> Surprisingly, even numbers were more likely to be dealt to the left side as compared to odd numbers (fixed effect estimate = 0.281, SE = 0.147), indicating a reliably reversed MARC effect,  $z = -1.91$ ,  $p = 0.028$ ,  $p_{pb} = 0.030$ .

For model comparisons, we fitted a null model including only an intercept on the first level, an additive model with value and parity as independent predictors, and a saturated model with main effects as well as the two-way interaction. In a first step, we compared the null model to the additive model. This comparison yielded a marginally significant effect in favor of the additive model  $\chi^2(2) = 5.42$ ,  $p = 0.067$ ,  $p_{pb} = 0.069$ , indicating that the two additional parameters did indeed add explanatory value. Further including the interaction effect, however, did not improve model fit significantly,  $\chi^2(1) = 0.01$ ,  $p = 0.941$ ,  $p_{pb} = 0.929$ .

### DISCUSSION

We investigated the effects of different characteristics of numbers (values of playing cards) on biases in fair distribution behavior. Indeed, we found evidence for such systematic biases in a free

choice experiment: Participants read out loud a rummy card's value and announced their spatial assignment to a leftward or rightward positioned player. Without applying explicit strategies, participants failed to distribute cards in a statistically fair way and assigned a mean benefit of two cards or 16 points to the right player. In line with recent findings from the linguistic markedness and spatial-numerical associations of response codes effects, we hypothesized such a pattern to be partly driven by odd and high numbers being associated with rightward oriented action codes.

In the following exploratory analyses, we aimed at dismantling SNARC and MARC-like effects on response decisions at an individual, trial-wise level. Indeed, we found some evidence for the regular SNARC effect, but the data also indicated a reversed MARC effect with odd numbers being more likely to be distributed to the right player and even numbers being more likely to be distributed to the left player. Although this latter finding certainly comes unexpected, several recent studies cast doubt on a stable left-right association of odd and even numbers. Rather, the direction of the MARC effect seems to depend on task rules, i.e., affirmative answers seem to be generally compatible with right response codes and might override the parity-driven code of an odd number (Cho and Proctor, 2007). Further, Nuerk et al. (2005) observed the MARC effect to be altered by stimulus and experimental settings: Whereas participants showed a usual MARC effect for number words when the experiment started with Arabic notation digits, this effect was reversed when the experiment started with dice-dot patterns. In light of the apparent similarity of dice patterns and the patterns printed on the playing cards of the current experiment (see Figure 1), one might speculate that such gambling-related stimuli might generally elicit a reversed linguistic markedness of parity; however, Chang and Gibson (2011) found a regular odd-even effect in Sudoku puzzles and future studies are needed to clarify these speculations and investigate the underlying mechanisms.

Such flexibility of the MARC effect further seems likely in light of various findings on flexible coding of the related SNARC effect. For instance, the SNARC effect is influenced by inter-individual characteristics such as finger counting habits (Fischer, 2008), cultural aspects such as reading direction (Shaki et al., 2009; Domahs et al., 2010) as well as sex (Bull et al., 2013) and age (Wood et al., 2008). The MARC effect, similarly, was recently found reversed for left-handers (Huber et al., 2014), which supports a body-specificity account (Casasanto, 2009) rather than a linguistic markedness account (Nuerk et al., 2004). Furthermore, the SNARC effect is also modulated by short-term, contextual factors such as recently encountered episodes (sequence effects: Pfister et al., 2013), number usage (number placement in text: Fischer et al., 2010; on a ruler vs. clock face: Bächtold et al., 1998) and current number range (Dehaene et al., 1993; Fias et al., 1996).

<sup>2</sup>Based on the comments of a reviewer, the model comparison was repeated using parametric bootstrapping with 1000 simulations, using the *PBmod-comp()* function of the R package *pbkrtest* (Halekoh and Højsgaard, 2014).

## GENERAL PLACEMENT PREFERENCES

Of course, the overall preference for the right card slot of our mostly right-handed participants also reminds of robust phenomena unrelated to the processing of numerical stimuli such as turning biases when confronted with a decision to take either a left or a right turn (Liederman and Kinsbourne, 1980; Güntürkün, 2003; cf. Shaki and Fischer, 2014, for the interplay of number processing and turning during walking). Furthermore, physical positioning was shown to produce more positive attitudes for rightward placed items (Nisbett and Wilson, 1977; Choi and Myer, 2012). *Vice versa*, positive abstract concepts were associated with right space for right-handed participants (Casasanto, 2009). In fact, a vast amount of marketing literature is concerned with devaluation of laterally placed items (Dittrich and Klauer, 2012), which is at times confounded with a desirable perception of magnitude (i.e., heaviness perception; Deng and Kahn, 2009) or automatic price and quality inferences (i.e., expensive and high-quality items on the right end; Valenzuela and Raghbir, 2009). For free choice actions, goal keepers were found more likely to dive to the right during shoot-outs and under pressure (Roskes et al., 2011; but see Price and Wolfers, 2014), which was taken to document approach motivation (Roskes et al., 2014).

## HANDEDNESS-DEPENDENT PLACEMENT PREFERENCES

Already for spontaneous turning biases, stronger right-sided head-turning was documented for right-handed than for left-handed participants (Ocklenburg and Güntürkün, 2009). Similarly, positive abstract concepts were associated with rightward space for right-handers, but left-handed participants with similar linguistic experience (i.e., use of metaphors) showed a reversed association of abstract concepts and space (Casasanto, 2009), suggesting that bodily experiences might shape valence-specific placement preferences. In a large Moroccan sample that exhibited strong taboos against the use of left hands, the implicit space-valence association was found effectively identical compared to a Spain sample (de la Fuente et al., 2014), but explicit measures (i.e., good-is-right rating and ratio of right/left-handers) were larger in the Arab population. Thus, handedness and according interactions with the external world appear to be valid candidates in explaining general and explicit spatial mappings of valence.

Given the data at hand, we cannot provide evidence for culture or hand-experience specific modulations. However, valence-space and value-space associations are not necessarily interchangeable, despite a possible positive connotation of playing cards or numbers in general. For mere numbers, reversing the polarity of a response side through response eccentricity did not affect spatial-numerical associations (Santiago and Lakens, 2014), suggesting that the link between numbers and space is not (exclusively) driven by their value-valence correspondence (i.e., polarity correspondence; Proctor and Cho, 2006). Another study even suggested magnitude to underlie spatial valence representations (Holmes and Lourenco, 2011). Furthermore, number-space associations are manifold regarding the number's features (see Patro et al., 2014, for a recent taxonomy proposal at an early age), and we next discuss the possible interpretation of SNARC and MARC effects in terms of linguistic markedness.

## LINGUISTIC MARKEDNESS IN NUMBER PROCESSING?

It is widely accepted that number processing includes a verbal component, as suggested by the triple-code model (Dehaene et al., 1993; Klein et al., 2014). Semantic features of the number (parity and magnitude) are activated automatically and can deteriorate unrelated task processing already in children of 10 years of age (Berch et al., 1999). As such, linguistic markedness of a verbal number-code, i.e., in form of the non-marked *even* parity feature, might facilitate equally non-marked responses, i.e., *right* actions (Nuerk et al., 2004). Arguably, in this experiment, the number of cards dealt to a player can be regarded an unspecific placement preference and explained a substantial proportion, but not all variance of differences in scores. Rather, the results from our exploratory analysis suggest that space-number associations further biased the distribution outcome, and that reversed space-parity associations supported but space-magnitude associations counteracted the fair distribution.

For linguistic influences in the SNARC effect, instead of assuming an oriented mental number line (i.e., Göbel et al., 2001), it is similarly possible that magnitude is coded by opposed small/large polar or linguistic representations (c.f. Nuerk et al., 2004; Proctor and Cho, 2006). Facilitated left/right responses can be accounted for by corresponding pairs of markedness: The adjectives *large* and *small* are lexical opposites with *large* as the non-marked adjective (Jakobson, 1931; see also: Lehrer, 2008). Similarly, the adjective *right* is linguistically non-marked (Zimmer, 1964), and the correspondence of both non-marked (i.e., *large* and *right*) and marked (i.e., *small* and *left*) pairs would lead to the SNARC effect. Homogenous marked and non-marked pairs should be responded to faster and they should more often be selected in a free choice paradigm. Consequently, with a decreasing marked property of *small*, the marked *left* response side was chosen less frequently. However, it is not clear how linguistic markedness can account for flexible magnitude-space and reversed parity-space associations; instead, a flexible, body-specific conceptual layer, i.e., in form of polarity or space, seems more likely. Obviously, participants were more cautious in distributing high-value (i.e., royal) cards more equally in order to distribute the cards fairly; nevertheless, magnitude-response correspondence, as indexed by the regular SNARC effect, could have effectively led to the observed right-bias.

Crucially, the interpretation of the SNARC effect in terms of polarity correspondence (Proctor and Cho, 2006) or verbal codes (Gevers et al., 2010) does not exclude the possibility of a visuo-spatial representation of magnitude. In line with the dual-coding framework of Paivio (1986), non-verbal and verbal representations can be processed referentially and activate each other. The observed SNARC effect in verbal and following motor responses can be attributed to such a referential activation. Possibly, a visuo-spatial representation was pronounced because our participants performed actual hand movements in a well-defined space, namely over a card-playing table.

We excluded two-digit and royal card stimuli from the mixed-effects SNARC and MARC models as too little is known about these indirectly magnitude-related stimuli at this time: Do they extend the mental number line similar to 0 (Pinhas and Tzelgov, 2012)? How are nominal two-digit numbers processed when part

of this specific number range (Dehaene et al., 1993; Nuerk and Willmes, 2005; Nuerk et al., 2014) and does the pictorial presentation, i.e., of a king vs. a jack, trigger marked representations other than the rule-based card value?

Notwithstanding these open issues, a range of recent papers addressed the linkages of brain mechanisms devoted to language and action, respectively, and elaborated these linkages in several frameworks to accommodate for SNARC and MARC effects (e.g., Pulvermüller, 2005; Barsalou, 2008; Fischer, 2012). In case of the SNARC effect, interestingly, language or number processing is *most likely* only indirectly associated with motor system activations through magnitude processing (Fias et al., 1996) and magnitude-related spatial codes (Gevers et al., 2006) or verbal codes (Gevers et al., 2010). Still, this indirect loop was demonstrated sufficient to modulate deliberate action selection (Daar and Pratt, 2008; Ruiz Fernández et al., 2011). In this experiment, we further show that this bias even transfers to a more natural card playing scenario and is able to interfere with a fair distribution task.

### FAIR DECISIONS IN CARD DISTRIBUTION

Although statistically the goal of fair distribution was not met, participants were mostly confident about their choices during debriefing and reported to have achieved the goal by deciding upon a subjective feeling of just distribution. This finding is in line with results on the egocentric fairness bias (Tanaka, 1999), stating that especially just world believers (Rubin and Peplau, 1975) consider their own behavior as fairer than other people's behavior. In relation to these findings, the perception of fairness might be considered biased by social demands (Blair, 2002), whereas actual fair behavior was counter-acted here by automaticity, i.e., number-space associations.

Several alternative explanations might also account for the observed general preference for the right player. In this regard, some limitations of the study have to be considered: Both the table coloring and the player characters were not counterbalanced and could have implied unidentified response tendencies<sup>3</sup>. The study sample was rather diverse regarding participants' age, sex, and handedness, which likely increased the variance of number-space associations. Future studies should more closely examine these characteristics' interactions with number-driven action decisions. By including the rummy card set, the stimuli used were, on one hand, of high external validity and allowed for instructing and investigating fair distribution behavior. On the other hand, the stimulus set by nature included two-digit and pictorial cards and thereby differs from previous studies. Nevertheless, we focused on single digits only in the mixed effects models analysis and thereby, the results of this analysis must be regarded exploratory and might underestimate the SNARC effect for the entire number range.

A closer look at single digits in the exploratory analysis pointed towards regular magnitude-space associations, but reversed parity-space associations. As such, automatic number magnitude processing emphasized a possible pre-existing preference bias by suggesting rightward (leftward) choices for high

(low) value cards, resulting in higher scores. Given the full standard rummy card set, a regular MARC effect would have further emphasized responses favoring the right player. Placement preferences were increasingly identified in the literature, and the same is true for number-space associations. In a natural setting, it is likely that both types of bias affect choices, and our analysis confirms this view by the combination of identity-unspecific results (number of cards) and number specific results (scores and single-digit decision outcomes).

In conclusion, the results of our study support current views of actions as being influenced by language processing. During card distribution and while aiming at a fair and equal distribution, the participants' choices were still affected by linguistic or conceptual features of actual rummy cards, namely digit parity and magnitude. A regular SNARC and a reversed MARC effect emerged and ultimately supported the overall preference of a right player avatar. The successful transfer of these effects to a more natural setting emphasizes the importance of further understanding the (neural) mechanisms behind indirectly and directly action-related linguistic and conceptual influences on number processing. Understanding these mechanisms will allow for identifying in which situations number associations can systematically bias behavior and, consequently, a better understanding will allow for countering these biases.

### AUTHOR CONTRIBUTIONS

PS and RP designed research; PS performed research; PS and RP analyzed data and wrote the paper.

### ACKNOWLEDGMENTS

We acknowledge support by Deutsche Forschungsgemeinschaft and Open Access Publishing Fund of Tübingen University. We are grateful to Julia Schönrock for casting and recruiting suitable Playmobil® characters for the study.

### REFERENCES

- Bächtold, D., Baumüller, M., and Brugger, P. (1998). Stimulus-response compatibility in representational space. *Neuropsychologia* 36, 731–735. doi: 10.1016/S0028-3932(98)00002-5
- Barsalou, L. (1999). Perceptions of perceptual symbols. *Behav. Brain Sci.* 22, 637–660. doi: 10.1017/S0140525X99532147
- Barsalou, L. W. (2008). Grounded cognition. *Annu. Rev. Psychol.* 59, 617–645. doi: 10.1146/annurev.psych.59.103006.093639
- Bates, D., Maechler, M., Bolker, B., and Walker, S. (2014). *Linear Mixed-Effects Models Using Eigen and S4*. R Package version 1.1-7. Available at: <http://CRAN.R-project.org/package=lme4>
- Berch, D. B., Foley, E. J., Hill, R. J., and Ryan, P. M. (1999). Extracting parity and magnitude from Arabic numerals: developmental changes in number processing and mental representation. *J. Exp. Child Psychol.* 74, 286–308. doi: 10.1006/jecp.1999.2518
- Blair, I. V. (2002). The malleability of automatic stereotypes and prejudice. *Personal. Soc. Psychol. Rev.* 6, 242–261. doi: 10.1207/S15327957PSPR0603\_8
- Bull, R., Cleland, A. A., and Mitchell, T. (2013). Sex differences in the spatial representation of number. *J. Exp. Psychol. Gen.* 142, 181–192. doi: 10.1037/a0028387
- Casasanto, D. (2009). Embodiment of abstract concepts: good and bad in right- and left-handers. *J. Exp. Psychol. Gen.* 138, 351–367. doi: 10.1037/a0015854
- Chang, H.-S., and Gibson, J. M. (2011). The odd-even effect in Sudoku puzzles: effects of working memory, aging, and experience. *Am. J. Psychol.* 124, 313–324. doi: 10.5406/amerjpsyc.124.3.0313
- Cho, Y. S., and Proctor, R. W. (2007). When is an odd number not odd? Influence of task rule on the MARC effect for numeric classification. *J. Exp. Psychol. Learn. Mem. Cogn.* 33, 832–842. doi: 10.1037/0278-7393.33.5.832

<sup>3</sup>For effects of color on cognition, see Elliot and Maier (2014). We thank a reviewer for drawing our attention to this point.

- Choi, J., and Myer, D. W. (2012). The effect of product positioning in a comparison table on consumers' evaluation of a sponsor. *Mark. Lett.* 23, 367–380. doi: 10.1007/s11002-012-9162-9
- Daar, M., and Pratt, J. (2008). Digits affect actions: the SNARC effect and response selection. *Cortex* 44, 400–405. doi: 10.1016/j.cortex.2007.12.003
- Dehaene, S., Bossini, S., and Giraux, P. (1993). The mental representation of parity and number magnitude. *J. Exp. Psychol. Gen.* 122, 371–396. doi: 10.1037/0096-3445.122.3.371
- de la Fuente, J., Casasanto, D., Román, A., and Santiago, J. (2014). Can culture influence body-specific associations between space and valence? *Cogn. Sci.* doi: 10.1111/cogs.12177 [Epub ahead of print].
- Deng, X., and Kahn, B. (2009). Is your product on the right side? The “location effect” on perceived product heaviness and package evaluation. *J. Mark. Res.* 46, 1–48. doi: 10.1509/jmkr.46.6.725
- Dittrich, K., and Klauer, K. C. (2012). Does ignoring lead to worse evaluations? A new explanation of the stimulus devaluation effect. *Cogn. Emot.* 26, 193–208. doi: 10.1080/02699931.2011.570313
- Domahs, F., Moeller, K., Huber, S., Willmes, K., and Nuerk, H.-C. (2010). Embodied numerosity: implicit hand-based representations influence symbolic number processing across cultures. *Cognition* 116, 251–266. doi: 10.1016/j.cognition.2010.05.007
- Eerland, A., Guadalupe, T. M., and Zwaan, R. A. (2011). Leaning to the left makes the Eiffel Tower seem smaller: posture-modulated estimation. *Psychol. Sci.* 22, 1511–1514. doi: 10.1177/0956797611420731
- Elliot, A. J., and Maier, M. A. (2014). Color psychology: effects of perceiving color on psychological functioning in humans. *Annu. Rev. Psychol.* 65, 95–120. doi: 10.1146/annurev-psych-010213-115035
- Fias, W., Brysbaert, M., Geypens, F., and D'Ydewalle, G. (1996). The importance of magnitude information in numerical processing: evidence from the SNARC effect. *Math. Cogn.* 2, 95–110. doi: 10.1080/135467996387552
- Fischer, M. H. (2012). A hierarchical view of grounded, embodied, and situated numerical cognition. *Cogn. Process.* 13(Suppl. 1), S161–S164. doi: 10.1007/s10339-012-0477-5
- Fischer, M. H. (2008). Finger counting habits modulate spatial-numerical associations. *Cortex* 44, 386–392. doi: 10.1016/j.cortex.2007.08.004
- Fischer, M. H. (2006). The future for SNARC could be stark. ... *Cortex* 42, 1066–1068. doi: 10.1016/S0010-9452(08)70218-1
- Fischer, M. H., Mills, R. A., and Shaki, S. (2010). How to cook a SNARC: number placement in text rapidly changes spatial-numerical associations. *Brain Cogn.* 72, 333–336. doi: 10.1016/j.bandc.2009.10.010
- Furnham, A., and Boo, H. C. (2011). A literature review of the anchoring effect. *J. Socio Econ.* 40, 35–42. doi: 10.1016/j.socsec.2010.10.008
- Gevers, W., Ratinckx, E., De Baene, W., and Fias, W. (2006). Further evidence that the SNARC effect is processed along a dual-route architecture: evidence from the lateralized readiness potential. *Exp. Psychol.* 53, 58–68. doi: 10.1027/1618-3169.53.1.58
- Gevers, W., Santens, S., Dhooge, E., Chen, Q., Van den Bossche, L., Fias, W., et al. (2010). Verbal-spatial and visuospatial coding of number-space interactions. *J. Exp. Psychol. Gen.* 139, 180–190. doi: 10.1037/a0017688
- Göbel, S., Walsh, V., and Rushworth, M. F. (2001). The mental number line and the human angular gyrus. *Neuroimage* 14, 1278–1289. doi: 10.1006/nimg.2001.0927
- Güntürkün, O. (2003). Human behaviour: adult persistence of head-turning asymmetry. *Nature* 421, 711. doi: 10.1038/421711a
- Halekoh, U., and Højsgaard, S. (2014). A Kenward-Roger approximation and parametric bootstrap methods for tests in linear mixed models—the R package pbrtest. *J. Stat. Softw.* 59, 1–30.
- Holmes, K. J., and Lourenco, S. F. (2011). Common spatial organization of number and emotional expression: a mental magnitude line. *Brain Cogn.* 77, 315–323. doi: 10.1016/j.bandc.2011.07.002
- Huber, S., Klein, E., Graf, M., Nuerk, H.-C., Moeller, K., and Willmes, K. (2014). Embodied markedness of parity? Examining handedness effects on parity judgments. *Psychol. Res.* doi: 10.1007/s00426-014-0626-9 [Epub ahead of print].
- Jakobson, R. (1931). “The structure of the Russian verb,” in *Russian and Slavic Grammar Studies*, eds L. Waugh and M. Halle (Berlin: Mouton), 1–14.
- Keehner, M., and Fischer, M. H. (2012). Unusual bodies, uncommon behaviors: individual and group differences in embodied cognition in spatial tasks. *Spat. Cogn. Comput.* 12, 71–82. doi: 10.1080/13875868.2012.659303
- Klein, E., Suchan, J., Moeller, K., Karnath, H.-O., Knops, A., Wood, G., et al. (2014). Considering structural connectivity in the triple code model of numerical cognition: differential connectivity for magnitude processing and arithmetic facts. *Brain Struct. Funct.* doi: 10.1007/s00429-014-0951-1 [Epub ahead of print].
- Lehrer, A. (2008). Markedness and antonymy. *J. Linguist.* 21, 397. doi: 10.1017/S00222670001032X
- Liederman, J., and Kinsbourne, M. (1980). The mechanism of neonatal rightward turning bias: a sensory or motor asymmetry? *Infant Behav. Dev.* 3, 223–238. doi: 10.1016/S0163-6383(80)80028-2
- Loetscher, T., Bockisch, C. J., Nicholls, M. E. R., and Brugger, P. (2010). Eye position predicts what number you have in mind. *Curr. Biol.* 20, R264–R265. doi: 10.1016/j.cub.2010.01.015
- Loetscher, T., Schwarz, U., Schubiger, M., and Brugger, P. (2008). Head turns bias the brain's internal random generator. *Curr. Biol.* 18, R60–R62. doi: 10.1016/j.cub.2007.11.015
- Mussweiler, T., and Englich, B. (2003). Adapting to the Euro: evidence from bias reduction. *J. Econ. Psychol.* 24, 285–292. doi: 10.1016/S0167-4870(03)00015-1
- Nisbett, R., and Wilson, T. (1977). Telling more than we can know: verbal reports on mental processes. *Psychol. Rev.* 84, 231–259. doi: 10.1037/0033-295X.84.3.231
- Northcraft, G. B., and Neale, M. A. (1987). Experts, amateurs, and real estate: an anchoring-and-adjustment perspective on property pricing decisions. *Organ. Behav. Hum. Decis. Process.* 39, 84–97. doi: 10.1016/0749-5978(87)90046-X
- Nuerk, H.-C., Iversen, W., and Willmes, K. (2004). Notational modulation of the SNARC and the MARC (linguistic markedness of response codes) effect. *Q. J. Exp. Psychol. A* 57, 835–863. doi: 10.1080/02724980343000512
- Nuerk, H.-C., Moeller, K., Klein, E., Willmes, K., and Fischer, M. H. (2011). Extending the mental number line: a review of multi-digit number processing. *Z. Psychol.* 219, 3–22. doi: 10.1027/2151-2604/a000041
- Nuerk, H.-C., Moeller, K., and Willmes, K. (2014). “Multi-digit number processing,” in *Oxford Handbook of Mathematical Cognition*, eds R. Cohen Kadosh and A. Dowker (Oxford: Oxford University Press), 1–20.
- Nuerk, H.-C., and Willmes, K. (2005). On the magnitude representations of two-digit numbers. *Psychol. Sci.* 47, 52–72.
- Nuerk, H.-C., Wood, G., and Willmes, K. (2005). The universal SNARC effect: the association between number magnitude and space is amodal. *Exp. Psychol.* 52, 187–194. doi: 10.1027/1618-3169.52.3.187
- Ocklenburg, S., and Güntürkün, O. (2009). Head-turning asymmetries during kissing and their association with lateral preference. *Laterality* 14, 79–85. doi: 10.1080/1357650080243689
- Paivio, A. (1986). *Mental Representations: A Dual-Coding Approach*. New York: Oxford University Press.
- Patro, K., Nuerk, H.-C., Cress, U., and Haman, M. (2014). How number-space relationships are assessed before formal schooling: a taxonomy proposal. *Front. Psychol.* 5:419. doi: 10.3389/fpsyg.2014.00419
- Pfister, R., and Janczyk, M. (2013). Confidence intervals for two sample means: calculation, interpretation, and a few simple rules. *Adv. Cogn. Psychol.* 9, 74–80.
- Pfister, R., Schroeder, P. A., and Kunde, W. (2013). SNARC struggles: instant control over spatial-numerical associations. *J. Exp. Psychol. Learn. Mem. Cogn.* 39, 1953–1958. doi: 10.1037/a0032991
- Pinhas, M., and Tzelgov, J. (2012). Expanding on the mental number line: zero is perceived as the “smallest.” *J. Exp. Psychol. Learn. Mem. Cogn.* 38, 1187–1205. doi: 10.1037/a0027390
- Porter, T. M. (1996). *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton, NJ: Princeton University Press.
- Price, J., and Wolfers, J. (2014). Right-oriented bias: a comment on roskes, sligte, shalvi, and de dreu (2011). *Psychol. Sci.* 25, 2109–2111. doi: 10.1177/0956797614536738
- Proctor, R. W., and Cho, Y. S. (2006). Polarity correspondence: a general principle for performance of speeded binary classification tasks. *Psychol. Bull.* 132, 416–442. doi: 10.1037/0033-2909.132.3.416
- Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nat. Rev. Neurosci.* 6, 576–582. doi: 10.1038/nrn1706
- Roskes, M., Slight, D., Shalvi, S., and De Dreu, C. K. W. (2014). Does approach motivation induce right-oriented bias? Reply to price and wolfers (2014). *Psychol. Sci.* 25, 2112–2115. doi: 10.1177/0956797614547919
- Roskes, M., Slight, D., Shalvi, S., and De Dreu, C. K. W. (2011). The right side? Under time pressure, approach motivation leads to right-oriented bias. *Psychol. Sci.* 22, 1403–1407. doi: 10.1177/0956797611418677

- Rubin, Z., and Peplau, L. A. (1975). Who Believes in a Just World? *J. Soc. Issues* 31, 65–89. doi: 10.1111/j.1540-4560.1975.tb00997.x
- Ruiz Fernández, S., Rahona, J. J., Hervás, G., Vázquez, C., and Ulrich, R. (2011). Number magnitude determines gaze direction: spatial-numerical association in a free-choice task. *Cortex* 47, 617–620. doi: 10.1016/j.cortex.2010.10.006
- Santiago, J., and Lakens, D. (2014). Can conceptual congruency effects between number, time, and space be accounted for by polarity correspondence? *Acta Psychol. (Amst.)* doi: 10.1016/j.actpsy.2014.09.016 [Epub ahead of print].
- Shaki, S., and Fischer, M. H. (2014). Random walks on the mental number line. *Exp. Brain Res.* 232, 43–49. doi: 10.1007/s00221-013-3718-7
- Shaki, S., Fischer, M. H., and Petrusic, W. M. (2009). Reading habits for both words and numbers contribute to the SNARC effect. *Psychon. Bull. Rev.* 16, 328–331. doi: 10.3758/PBR.16.2.328
- Tanaka, K. (1999). Judgments of fairness by just world believers. *J. Soc. Psychol.* 139, 631–638. doi: 10.1080/00224549909598423
- Tschentscher, N., Hauk, O., Fischer, M. H., and Pulvermüller, F. (2012). You can count on the motor cortex: finger counting habits modulate motor cortex activation evoked by numbers. *Neuroimage* 59, 3139–3148. doi: 10.1016/j.neuroimage.2011.11.037
- Tversky, A., and Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science* 185, 1124–1131. doi: 10.1126/science.185.4157.1124
- Valenzuela, A., and Raghurir, P. (2009). Are top-bottom inferences conscious and left-right inferences automatic? implications for shelf space positions. *HEC Paris INSEAD Fontainebleau*. 19, 185–196.
- Wood, G., Willmes, K., Nuerk, H.-C., and Fischer, R. (2008). On the cognitive link between space and number: a meta-analysis of the SNARC effect. *Psychol. Sci. Q.* 50, 489–525.
- Zimmer, K. E. (1964). Affixal negation in English and other languages: an investigation of restricted productivity. *Word* 20(2), Monograph No. 5.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 22 November 2014; accepted: 16 February 2015; published online: 20 March 2015.

Citation: Schroeder PA and Pfister R (2015) Arbitrary numbers counter fair decisions: trails of markedness in card distribution. *Front. Psychol.* 6:240. doi: 10.3389/fpsyg.2015.00240

This article was submitted to *Developmental Psychology*, a section of the journal *Frontiers in Psychology*.

Copyright © 2015 Schroeder and Pfister. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# On the relation between grammatical number and cardinal numbers in development

Barbara W. Sarnecka \*

Department of Cognitive Sciences, University of California at Irvine, Irvine, CA, USA

**Edited by:**

Ann Dowker, University of Oxford, UK

**Reviewed by:**

Emily Mather, University of Hull, UK  
Ruth Ford, Anglia Ruskin University, UK

**\*Correspondence:**

Barbara W. Sarnecka, Department of Cognitive Sciences, University of California at Irvine, 3151 Social Sciences Plaza A, Irvine, CA 92697-5100, USA  
e-mail: sarnecka@uci.edu

This mini-review focuses on the question of how the grammatical number system of a child's language may help the child learn the meanings of cardinal number words (e.g., "one" and "two"). Evidence from young children learning English, Russian, Japanese, Mandarin, Slovenian, or Saudi Arabic suggests that trajectories of number-word learning differ for children learning different languages. Children learning English, which distinguishes between singular and plural, seem to learn the meaning of the cardinal number "one" earlier than children learning Japanese or Mandarin, which have very little singular/plural marking. Similarly, children whose languages have a singular/dual/plural system (Slovenian and Saudi Arabic) learn the meaning of "two" earlier than English-speaking children. This relation between grammatical and cardinal number may shed light on how humans acquire cardinal-number concepts. There is an ongoing debate about whether mental symbols for small cardinalities (concepts for "oneness," "twoness," etc.) are innate or learned. Although an effect of grammatical number on number-word learning does not rule out nativist accounts, it seems more consistent with constructivist accounts, which portray the number-learning process as one that requires significant conceptual change.

**Keywords:** cardinal, counting, language development, number, plural, grammatical number

There are different ways to convey numerical information in language. Suppose you and I meet for the first time, and you wonder whether I have children. (Of course you are too polite to ask.) During our conversation, I say, "I thought that as a developmental psychologist, I would find it easy to be a parent, but I don't." Now you know that I have at least one child. If I say, "I came to this conference to get away from my kids," you know that I have two or more children, because the English word *kids* is plural, and must refer to sets of two or more. Finally, if I say, "My kids can't stop arguing; they both want the last word," you know that I have exactly two children, because the English word *both* always refers to sets of exactly two. (A rare example of dual marking in English.) Alternatively, you might simply ask whether I have children, and I might say, "Yes. I have two boys."

As this example demonstrates, numerical information can be communicated via cardinal number words ("one," "two," "three," etc.), but it can also be communicated via grammatical morphology, such as the *s* on the English word *kids*. English is a singular/plural language, meaning that it marks the difference between sets of one and sets of two or more. But not all languages do this. Numeral classifier languages such as Japanese and Mandarin have very little singular/plural marking (Downing, 1996). In these languages, saying "I have kid(s)" is like saying in English, "I am a parent." It conveys no information at all about *how many* kids you have. Still other languages have singular/dual/plural marking systems, which pick out sets of one,

sets of two, and sets of three or more. In these languages, dual-marked noun phrases refer to sets of exactly two, similar to the English word *both*. A few languages go even further, marking singular/dual/trial/plural for sets of one, two, three, and four or more, respectively, or marking singular/dual/paucal/plural where paucal marking picks out small sets (something like the English phrase "a handful") and plural marking picks out larger sets (Corbett, 2000).

This mini-review focuses on the question of how of these two systems (grammatical number and cardinal numbers) may be related in development. There is some evidence that the grammatical number marking system of the language a child is learning may influence that child's learning of the cardinal number system. Because cardinal number systems are functionally identical across languages while grammatical number systems differ, we can look at differences in children's learning of cardinal numbers, and see if that learning bears the signatures of particular languages' grammatical number systems.

When we do this, we find evidence that indeed, a language's grammatical number system does seem to influence children's learning of cardinal number words in that language. Children learning a language as English, which pervasively marks singular/plural, seem to learn the meaning of the number "one" earlier than children whose languages do not mark singular/plural, such as Japanese (Sarnecka et al., 2007). Similarly, children whose languages have a singular/dual/plural system (Slovenian and Saudi Arabic) appear to learn the meaning of

“two” earlier than English-speaking children (Almoammer et al., 2013).

This is interesting, not because it tells us anything about how adult number concepts in any language, but because it may shed some light on how number concepts are acquired. There is an ongoing debate about whether mental symbols for small cardinalities (concepts for oneness, twoness, threeness, and the like) are innate or learned. Some proposals argue that these concepts are innate and shared with other animals (e.g., Gelman and Gallistel, 1978, 2004; Gelman and Butterworth, 2005; Butterworth et al., 2008). On these accounts, the challenge for the child learning language may just be to identify the words (i.e., cardinal number words) that match her innate concepts of oneness, twoness, threeness, etc.

On the other side of the debate, it is argued that humans are not born with concepts of oneness, twoness, threeness, etc., but must construct them (Le Corre and Carey, 2007; Carey, 2009). People in numerate societies construct these concepts during early childhood, in the course of learning the meanings for the cardinal number words “one,” “two,” “three,” and eventually the properties of the cardinal number system: that each number has a successor, that all sets of the same number can be put into one-to-one correspondence with each other, etc. (Izard et al., 2008, 2014; Sarnecka and Carey, 2008; Carey, 2009; Sarnecka and Wright, 2013; Sarnecka et al., in press).

## THE QUESTION

The question of how grammatical number might be related to cardinal number began with an observation about trajectories of number-word learning in English. In the early 1990s, Wynn (1990, 1992) first reported that children learn the meanings of cardinal number words one at a time and in order. Wynn showed this using the “Give-N” or “Give-a-number” task, in which she asked children to give her a certain number of items (e.g., “Give me one fish”; “Give me three fish,” etc.). She found that children’s performance moved through a predictable series of levels.

At the earliest (“pre-number-knower”) level, children do not distinguish among the different number words. Pre-number knowers might give one object for every number requested, or they might give a handful of objects for every number, but they show no sign of knowing the exact meaning of any number word. At the next level (called the “one-knower” level), children know that “one” means 1. On the Give-N task, one-knowers give exactly one object when asked for “one,” and they give two or more objects when asked for any other number. After this comes the “two-knower” level, where children give one object for “one,” and two objects for “two,” but do not reliably produce larger sets. This is followed by a “three-knower” level and (although Wynn didn’t find it because she never asked children for four objects) a “four-knower” level. After the four-knower level, children seem to learn the meanings of the higher cardinal number words in a different way—inferring their meanings from their place in the counting list rather than learning them individually as they did with the small numbers (Carey, 2009). Children who have done this (i.e., who have figured out how the counting system represents cardinal numbers) are called “Cardinal-principle knowers.”

The age at which children master these knower levels differs from one child to another, but in the most commonly studied population (English-speaking children from relatively privileged socioeconomic backgrounds), children typically reach the “one-knower” level some time during their second or third year (i.e., between 24 and 47 months old) and reach the final, “cardinal-principle-knower” level about 1 year later, between about 34 and 51 months (Sarnecka et al., in press).

As a graduate student reading Wynn’s work in the late 1990s, I noticed a parallel between children’s number-word learning and grammatical number systems. Both follow a rigid hierarchy: a child who understands “two” always understands “one” as well, just as a language that marks dual always marks singular as well. There do not seem to be children who understand “three” but *not* “one” and “two,” just as there are no languages that grammatically mark trial but *not* singular and dual. In a way, pre-number-knowers are like speakers of numeral classifier languages (e.g., Japanese); one-knowers are like speakers of singular/plural languages (e.g., English); and two-knowers were like speakers of singular/dual/plural languages (e.g., Slovenian).

A striking feature of number-word learning in English is the really long one-knower level. Wynn (1992) reported that children seemed to spend many months at the one-knower level—much longer than they spent as two-knowers or three-knowers. Why should that be the case? One possible explanation is that because English is a singular/plural language, English-speaking children must pay special attention to the distinction between one and other set sizes. English-speaking children show understanding of singular/plural marking between 20 and 24 months of age (Kouider et al., 2006); it is possible that this knowledge helps children learn the meaning of “one” sooner than they would if their language did not distinguish singular from plural. This explanation can be tested by comparing number-word learning in English to number-word learning in Japanese, which generally does not distinguish singular from plural.

A different possibility is that “one” is learned earlier than “two” simply because “one” is much more frequent in everyday speech. Across languages, “one” is more frequent than “two”; “two” is more frequent than “three,” and so on (Dehaene and Mehler, 1992). The frequency of “one” is particularly high in English, where it appears not only in counting, but also in deictic and anaphoric contexts (e.g., “Look at that one” or, “I’m making sandwiches—do you want one?”) This explanation can be tested by comparing English-speaking children’s number-word learning to that of children speaking Russian, a singular/plural language where the cardinal number “one” does not appear in non-numeric contexts.

## THE EVIDENCE

My collaborators and I administered Wynn’s Give-a-number task, as well as a counting task, to young children living in Ann Arbor, MI, USA; St. Petersburg, Russia, and Kobe, Japan (Sarnecka et al., 2007). Children in each group ranged in age from 2 years, 9 months to 3 years, 6 months, and the mean age for each group was 3 years, 2 months.

We found that more English- and Russian-speakers knew the meaning of “one” than did their Japanese counterparts,

supporting the idea that speaking a language with singular/plural marking helps children learn the meaning of “one.” Comparing English to Russian, we found that Russian-speakers were actually more likely to know “one” than English speakers, even though the Russian word for “one” appears less frequently in everyday speech than the English word “one.” Thus, the data did not support the idea that the overall high frequency of “one” relative to other numbers causes English-speaking children to reach the one-knower level sooner. Rather, it seems to be the presence of singular/plural marking in the language that makes the difference.

One question that arose about these findings was whether Japanese was the best choice to represent non-singular/plural marking languages. Number-word learning in Japanese is potentially complicated by the presence of two count lists, which sound nothing at all alike. (One of the lists begins *ichi, ni, san, shi, go...* the other begins *hitotsu, futatsu, mitsu, yotsu, itsutsu...*) Both of the lists are commonly used for numbers up to 10 (although only the *ichi, ni, san* list is used for numbers above 10), so it is reasonable to ask whether Japanese children might take longer to learn the number-word meanings, just because the input they receive for each number is potentially divided between two different word forms.

We addressed this question in the 2007 paper by arguing that Russian-speaking children also have to deal with different word forms, as numbers are declined for gender and case. For example, the word *one* in Russian may take any of the following forms: *odin, odna, odno, odni, odnu, odnovo, odnikh, odnoy, odnom, odnomu, odnim, odnimi*. But this argument is not wholly convincing, first because these forms of *one* are not as different from each other as *hitotsu* and *ichi*, and second because when people actually count in Russian, the number words are usually in the nominative case, so the count list sounds the same every time. Japanese, on the other hand, actually has two different counting lists, which could be a serious confound. So it is important to note that the finding of children learning “one” later in a non-singular/plural language has not only been replicated in Japanese (Barner et al., 2009b) but is also found in Mandarin, which very sensibly has only one count list (Li et al., 2003).

Further evidence for a link between grammatical number and cardinal number-word learning has recently come from a study with young speakers of two languages with singular/dual/plural systems: Slovenian and Saudi Arabic (Almoammer et al., 2013). The study tested 2- to 4-year-old children in Slovenian, and 3- and 4-year-old children in Arabic. Significantly more children knew the meaning of “two” in the dual-marking languages than in age-matched groups of English speakers. Slovenian children learned “two” sooner than English-speaking children despite not being able to count as well as the English speakers, which is surprising because counting ability would seem to indicate experience with numbers. (No counting data were available for the Saudi Arabic-speaking children.) In both Slovenian and Saudi Arabic, children’s understanding of the grammatical dual forms was correlated with their knowledge of the cardinal number “two.”

Moreover, just as English-speaking children seem to spend a long time at the one-knower level, so do Slovenian-speaking children spend a long time at the two-knower level. Although they learn “two” earlier, they stay at the two-knower level for

longer, taking more time to learn “three” and higher numbers than children in the other language environments studied. This connection between grammatical dual marking and learning “two” is interesting because it shows that the meaning of “two” doesn’t follow automatically from “one,” but requires additional inference, for which dual-marking languages provide additional evidence. This pattern is consistent with Carey’s (2009) account, in which the meanings of “one” through “four” are learned individually, whereas the meanings of the higher numbers are learned as a group, when the child comes to understand the cardinal principle.

At least one qualification to these findings should be noted. In our original paper, we speculated that children learning singular/plural languages like English may initially understand “one” as meaning *singular* as opposed to *plural* (Sarnecka et al., 2007). As an example, we suggested that children may treat “one” like the indefinite article “a(n).” (In fact, the number “one” and the indefinite article were originally the same word in English, as they are today in languages such as Spanish and French.)

However, one study compared English-speaking children’s use of “one” and “a(n),” and found that children sometimes treat them differently. Children were shown a plate with two apples on it, and were asked either, “Is there *an apple* on the plate?” or “Is there *one apple* on the plate?” (Barner et al., 2009a). Children generally agreed with the statement that there was “an apple” on the plate, but disagreed with the statement that there was “one apple,” indicating that they treated the number “one” as upper-bounded (i.e., more than one is not one), but did not treat the word “a(n)” that way. Thus, although grammatical number helps children learn the meaning of “one,” they do not treat the words as identical.

## CONCLUSION

It does appear that the child’s learning of cardinal numbers is affected by the grammatical number system of his or her native language. Children whose languages mark singular/plural learn the cardinal meaning of the counting word “one” sooner than children whose languages do not mark the singular/plural distinction. Similarly, children whose languages distinguish dual from both singular and plural seem to learn “two” earlier than children in other language environments.

Even more interesting, perhaps, is the slight delay that children seem to experience in learning the first number *not* grammatically marked by their language. That is, children speaking singular/plural languages not only learn “one” a little sooner, but also seem to stay at the one-knower stage a bit later than children speaking other languages. Similarly, children whose languages include dual marking not only learn “two” earlier, but also seem to linger at the two-knower level longer than children in other language environments.

This suggests that the process of learning numbers that are grammatically marked (i.e., “one” for speakers of singular/plural languages; “one” and “two” for speakers of singular/dual/plural languages) may differ from the process of learning numbers that are not so marked. Children may use different sources of information to learn the meanings of grammatically marked vs. unmarked numbers. When the information from grammar runs

out (e.g., when English speakers move on to learning “two” or Slovenian speakers to learning “three”), children must rely on some other source of information to figure out the next number word. This results in a slight delay in learning, relative to speakers of languages such as Japanese where all numbers are learned without the help of grammatical number marking<sup>1</sup>.

If number-word learning is affected by the child’s language environment, what if anything does that tell us about the innateness of number concepts? On balance, this evidence seems most compatible with constructivist views, because it implies that number-word learning requires significant conceptual change.

When a child’s language environment highlights certain numerical distinctions (i.e., one/more than one, or one/two/more than two), these distinctions become more salient to the child, and therefore more available as candidate meanings for counting words, speeding the number-acquisition process. Perhaps having to distinguish between individuals and sets (or between individuals, pairs, and larger sets), speeds number learning by making concepts such as *individual*, *pair*, and *set* available as candidate meanings for cardinal number words.

Similarly, children slow down a bit when they encounter the first number whose meaning is not grammatically marked. This implies that children learn grammatically marked and unmarked numbers by different processes, which is also seems more consistent with a constructivist than a nativist framework.

Of course, it is possible to hold a nativist position and still allow that grammatical distinctions can help children map counting words to innate number concepts. But overall, these effects of environment on learning seem to support constructivist accounts, where children build concepts of oneness, twoness, threeness, etc. based on the particular evidence they have available. When the grammatical number system of a language highlights different numerical distinctions, trajectories of cardinal number learning differ in systematic and predictable ways. This implies that becoming numerate involves something more than simply a matching a verbal counting list to an innate, non-verbal counting list. Numerate children, it implies, are made and not born.

## REFERENCES

- Almoammer, A., Sullivan, J., Donlan, C., Marušič, F., Žaucer, R., O’Donnell, T., et al. (2013). Grammatical morphology as a source of early number word meanings. *Proc. Natl. Acad. Sci. U.S.A.* 110, 18448–18453. doi: 10.1073/pnas.1313652110
- Barner, D., Chow, K., and Yang, S.-J. (2009a). Finding one’s meaning: a test of the relation between quantifiers and integers in language development. *Cogn. Psychol.* 58, 195–219. doi: 10.1016/j.cogpsych.2008.07.001
- Barner, D., Libenson, A., Cheung, P., and Takasaki, M. (2009b). Cross-linguistic relations between quantifiers and numerals in language acquisition: evidence from Japanese. *J. Exp. Child Psychol.* 103, 421–440. doi: 10.1016/j.jecp.2008.12.001
- Butterworth, B., Reeve, R., Reynolds, F., and Lloyd, D. (2008). Numerical thought with and without words: evidence from indigenous Australian children. *Proc. Natl. Acad. Sci. U.S.A.* 105, 13179–13184. doi: 10.1073/pnas.0806045105
- Carey, S. (2009). *The Origin of Concepts*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780195367638.001.0001
- Corbett, G. G. (2000). *Number. Cambridge Textbooks in Linguistics*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139164344
- Dehaene, S., and Mehler, J. (1992). Cross-linguistic regularities in the frequency of number words. *Cognition* 43, 1–29. doi: 10.1016/0010-0277(92)90030-L
- Downing, P. A. (1996). *Numerical Classifier Systems: The Case of Japanese*, Vol. 4. John Benjamins Publishing. Retrieved from <http://books.google.com/books?hl=en&lr=&id=McpBAAAQBAJ&oi=fnd&pg=PR1&dq=Downing+Japanese&ots=Uzt0exIXQ&sig=eLuUD5FrjeLC4bVU0eoiDy3YE>
- Gelman, R., and Butterworth, B. (2005). Number and language: how are they related? *Trends Cogn. Sci.* 9, 6–10. doi: 10.1016/j.tics.2004.11.004
- Gelman, R., and Gallistel, C. R. (1978). *The Child’s Understanding of Number*. Cambridge, MA: Harvard University Press.
- Gelman, R., and Gallistel, C. R. (2004). Language and the origin of numerical concepts. *Science* 306, 441–443. doi: 10.1126/science.1105144
- Izard, V., Pica, P., Spelke, E. S., and Dehaene, S. (2008). Exact equality and successor function: two key concepts on the path towards understanding exact numbers. *Philos. Psychol.* 21, 491–505. doi: 10.1080/09515080802825354
- Izard, V., Streri, A., and Spelke, E. S. (2014). Toward exact number: young children use one-to-one correspondence to measure set identity but not numerical equality. *Cogn. Psychol.* 72, 27–53. doi: 10.1016/j.cogpsych.2014.01.004
- Kouider, S., Halberda, J., Wood, J., and Carey, S. (2006). Acquisition of English number marking: the singular–plural distinction. *Lang. Learn. Dev.* 2, 1–25. doi: 10.1207/s15473341l1d0201\_1
- Le Corre, M., and Carey, S. (2007). One, two, three, four, nothing more: an investigation of the conceptual sources of the verbal counting principles. *Cognition* 105, 395–438. doi: 10.1016/j.cognition.2006.10.005
- Li, P., Le Corre, M., Shui, R., Jia, G., and Carey, S. (2003). Effects of plural syntax on number word learning: a cross-linguistic study. *Presented at the 28th Boston University Conference on Language Development*, Boston, MA.
- Sarnecka, B. W., and Carey, S. (2008). How counting represents number: what children must learn and when they learn it. *Cognition* 108, 662–674. doi: 10.1016/j.cognition.2008.05.007
- Sarnecka, B. W., Goldman, M. C., and Slusser, E. B. (in press). “How counting leads to children’s first representations of large, exact numbers,” in *Oxford Handbook of Numerical Cognition*, eds R. Cohen Kadosh and A. Dowker (NY: Oxford University Press).
- Sarnecka, B. W., Kamenskaya, V. G., Yamana, Y., Ogura, T., and Yudovina, Y. B. (2007). From grammatical number to exact numbers: early meanings of “one”, “two”, and “three” in English, Russian, and Japanese. *Cogn. Psychol.* 55, 136–168. doi: 10.1016/j.cogpsych.2006.09.001
- Sarnecka, B. W., and Wright, C. E. (2013). The idea of an exact number: children’s understanding of cardinality and equinumerosity. *Cogn. Sci.* 37, 1493–1506. doi: 10.1111/cogs.12043
- Wynn, K. (1990). Children’s understanding of counting. *Cognition* 36, 155–193. doi: 10.1016/0010-0277(90)90003-3
- Wynn, K. (1992). Children’s acquisition of the number words and the counting system. *Cogn. Psychol.* 24, 220–251. doi: 10.1016/0010-0285(92)90008-P

**Conflict of Interest Statement:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 16 July 2014; accepted: 17 September 2014; published online: 09 October 2014.

Citation: Sarnecka BW (2014) On the relation between grammatical number and cardinal numbers in development. *Front. Psychol.* 5:1132. doi: 10.3389/fpsyg.2014.01132 This article was submitted to *Developmental Psychology*, a section of the journal *Frontiers in Psychology*.

Copyright © 2014 Sarnecka. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

<sup>1</sup>Thanks to Emily Mather for pointing out the potential significance of the delay in learning the first grammatically unmarked number.

# Word problems: a review of linguistic and numerical factors contributing to their difficulty

Gabriella Daroczy<sup>1,2</sup>, Magdalena Wolska<sup>2</sup>, Walt Detmar Meurers<sup>2,3</sup> and Hans-Christoph Nuerk<sup>1,2,4\*</sup>

<sup>1</sup> Diagnostics and Cognitive Neuropsychology, Department of Psychology, Eberhard Karls Universität Tübingen, Tübingen, Germany, <sup>2</sup> LEAD Graduate School, Eberhard Karls Universität Tübingen, Tübingen, Germany, <sup>3</sup> Department of Linguistics, Eberhard Karls Universität Tübingen, Tübingen, Germany, <sup>4</sup> Knowledge Media Research Center, Tübingen, Germany

## OPEN ACCESS

### Edited by:

Yvette Renee Harris,  
Miami University, USA

### Reviewed by:

Catherine Thevenot,  
University of Geneva, Switzerland  
Lieven Verschaffel,  
University of Leuven, Belgium

### \*Correspondence:

Hans-Christoph Nuerk,  
Diagnostics and Cognitive  
Neuropsychology, Department of  
Psychology, Eberhard Karls  
Universität Tübingen,  
Schleichstrasse 4,  
72074 Tübingen, Germany  
hc.nuerk@uni-tuebingen.de

### Specialty section:

This article was submitted to  
*Developmental Psychology*, a section  
of the journal *Frontiers in Psychology*

**Received:** 09 December 2014

**Paper pending published:**

21 December 2014

**Accepted:** 11 March 2015

**Published:** 01 April 2015

### Citation:

Daroczy G, Wolska M, Meurers WD  
and Nuerk H-C (2015) Word  
problems: a review of linguistic and  
numerical factors contributing to their  
difficulty.  
*Front. Psychol.* 6:348.  
doi: 10.3389/fpsyg.2015.00348

Word problems (WPs) belong to the most difficult and complex problem types that pupils encounter during their elementary-level mathematical development. In the classroom setting, they are often viewed as merely arithmetic tasks; however, recent research shows that a number of linguistic verbal components not directly related to arithmetic contribute greatly to their difficulty. In this review, we will distinguish three components of WP difficulty: (i) the linguistic complexity of the problem text itself, (ii) the numerical complexity of the arithmetic problem, and (iii) the relation between the linguistic and numerical complexity of a problem. We will discuss the impact of each of these factors on WP difficulty and motivate the need for a high degree of control in stimuli design for experiments that manipulate WP difficulty for a given age group.

**Keywords:** word problems, linguistics complexity, numerical complexity, text properties, difficulty

## Word Problems

### Introduction

Word problems (WPs) are part of the school curriculum and are taught at all levels of education. In WPs, relevant information is presented in the form of a short narrative rather than in mathematical notation (Verschaffel et al., 2000). Sometimes WPs specifically encode a quantitative relation between objects (Boonen et al., 2013). Many children from kindergarten through adulthood have severe difficulties in solving WPs (Nesher and Teubal, 1975; Riley et al., 1983; Lewis and Mayer, 1987; Hegarty et al., 1992; Verschaffel et al., 1992). Both linguistic and numerical complexity contributes to the difficulty in solving WPs. However, researchers have so far often focused on the one or the other aspect, depending on which field they come from. Even within the respective fields, linguistics, and numerical cognition, some aspects have been studied extensively, while others have been (strangely) neglected. For instance, we will see that semantics and discourse structures have been frequently studied in the context of WP complexity, but systematic syntactic manipulations are scarce. As regards numerical cognition, number properties like parity and magnitude as well as the type of mathematical reasoning have often been studied, but the type and the form of operations (e.g., carry-over effects) have not been investigated thoroughly in WPs, although they play an important role in current numerical cognition research (Moeller et al., 2011; Nuerk et al., 2011, 2015).

In this review, as researchers from the field of linguistics and the field of numerical cognition we have collaborated to provide a systematic overview of linguistic and numerical aspects relevant to solving WPs as well as their interaction. To capture a broad range of relevant facets

in the review, we extended our view of the relevant literature with systematic keyword searches in several databases (Web of Science, Ebsco, Google Scholar, ScienceDirect) including the following terms: WPs, story problems in combination with situational model, performance, consistency hypothesis, language processing, relational terminology, semantic influence, rewording, semantic cues, number size and type, working memory, text comprehension, computational errors, operations, position of unknown. In **Table 1**, we present selected linguistic, mathematical and general factors investigated in previous studies.

### Individual Differences and Social Factors

Individual differences and social factors must be also considered in WP research (Fite, 2002). For example, in the PISA studies—often measured with WPs—, mathematics literacy is a commonly used notion (Stacey, 2012). “It is defined as an individual’s capacity to identify and understand the role that mathematics plays in the world, to make well-founded judgments and to use and engage with mathematics in ways that meet the needs of that” (OECD, 2010). Unsuccessful WP solvers can experience negative social health and life outcome (Schley and Fujita, 2014). Even beyond social consequences, numerous studies focused on individual differences and group differences, such

as students with and without learning disabilities (Kingsdorf and Krawec, 2014), and children with and without developmental disabilities (Neef et al., 2003). Hegarty et al. (1995) distinguished domain-specific strategies that successful and unsuccessful problem solvers develop with practice and how these strategies account for individual differences in performance. Different students – e.g., individuals with calculation difficulty, or WP difficulty (Powell and Fuchs, 2014) – may struggle with different types of WPs. Besides domain-general capabilities like IQ, the role of domain specific knowledge and processes were investigated to get a complete account of problem solving, basic cognitive abilities; visual, reading skills, mathematical skills, and metacognitive abilities involved in the solution process. For example Boonen et al. (2014) and Oostermeijer et al. (2014) explored the role of spatial ability and reading comprehension in WP solving, since good WP solvers do not select numbers and relational keywords but create a visual representation (Boonen et al., 2013).

*Social factors* like schooling, teachers and peers also deserve consideration because the way of responding [e.g., De Corte et al. (1988)], the scoring criteria, the presence of illustrations next to the text [e.g., Berends and van Lieshout (2009)], or solution models used by the teachers influence WP performance considerably. School WPs also support stereotypical thinking: WPs do not

**TABLE 1 | Selected linguistic, mathematical, and general factors investigated in previous studies.**

Linguistic factors	Mathematical factors	General factors
<b>Structure</b> Structural complexity of basic quantitative properties (e.g., Number of letters, word and sentence length, proportion of complex words) (Searle et al., 1974; Neshet, 1976; Lepik, 1990) Vocabulary level (e.g., polysemous words, prepositional phrases, passive voice, clause structure; Spanos et al., 1988; Abedi et al., 1997; Abedi and Lord, 2001; Shaftel et al., 2006; Martiniello, 2008) Question wording/placing (Cummins et al., 1988)	<b>Property of numbers</b> Single digit (Lean et al., 1990) Multi digit (Haghverdi et al., 2012) Type of number [e.g., fraction: (Raduan, 2010), decimal number: problem size, role of number (De Corte et al., 1988)] Number magnitude [e.g., range of number smaller than 100: (Neshet, 1976)]	<b>Skills and social aspects</b> Social consequences (Schley and Fujita, 2014) Learning disabilities (Kingsdorf and Krawec, 2014) Successful/unsuccessful problem solvers (Hegarty et al., 1995) Calculation/word problem (WP) difficulties (Powell and Fuchs, 2014) Children/adults (De Corte et al., 1990; Hegarty et al., 1992)
	<b>Required operation</b> Addition subtraction (De Corte and Verschaffel, 1987) Multiplication division (De Corte et al., 1988) Given number (De Corte et al., 1990; Vicente et al., 2007)	<b>Categorization</b> Semantic structure of arithmetic WPs (Riley et al., 1983) Algebra textbook frequency (Mayer, 1981) Standard/non-standard WP (Jimenez and Verschaffel, 2014)
<b>Semantics</b> Linguistics verbal cues (van der Schoot et al., 2009) Phrasing in cue words (LeBlanc and Weber-Russell, 1996) Conceptual rewording (Vicente et al., 2007) Semantic/Object relation (?) Presence of distractor (Muth, 1992)	<b>Mathematical solution strategy</b> Counting from larger number (De Corte and Verschaffel, 1987) Position of the unknown (Garcia et al., 2006) Arithmetic fact retrieval (Orrantia et al., 2010) Number combination (Fuchs et al., 2009) Situation/Mental arithmetic strategy (Brissiaud and Sander, 2010)	<b>Solution strategies</b> Algebra WP/arithmetic WP (Koedinger and Nathan, 2004) WP solving theory-models (Kintsch and Greeno, 1985) Translation strategies (Hegarty et al., 1995) Spatial/visual representation (Boonen et al., 2013) Situation model (Thevenot et al., 2007)
	<b>Relevance of information</b> (Terao et al., 2004) <b>Numerical processes and representation</b> (MacGregor and Price, 1999; Goebel et al., 2014)	<b>Other aspects</b> Pedagogical factors (Lean et al., 1990) Socio-mathematics (Reusser, 1988) Stereotypes thinking (Yee and Lee, 1997) Real-word knowledge (Verschaffel et al., 1997) Response mode (De Corte et al., 1988)
	Consistency effect (Lewis and Mayer, 1987) e.g., revisited: (Pape, 2003)	Computer tutors (Nathan et al., 1992) Computer simulation (Dellarosa, 1986)
Basic linguistics influence on numerical cognition (Lachmair et al., 2014)		
	Working memory (Swanson et al., 1993)	

resemble problems in real-world situations (Yee and Lee, 1997). In addition, there is a strong tendency among both students and teachers to exclude real-world knowledge from their WP solution (Verschaffel et al., 1997), which is consistent with the observation that the problem solving process is also influenced by social cognitive and epistemic behavior settings (Reusser, 1988). Linguistic and pedagogical factors also affect children's understanding of arithmetic WPs (Lean et al., 1990). Students' beliefs about what doing and knowing mathematics means are rather different from the ideals (Jimenez and Verschaffel, 2014) and shaped by "socio-mathematical norms." Resulting differences in motivation seem to influence the strategies used to solve WPs (Gasco and Villarroel, 2014). In sum, both individual differences and social factors contribute to WP performance and deserve consideration.

### Subcategories of Word Problems and Solution Strategies

Several different types of WPs—e.g., in the underlying mathematical structure or solvability—are often presented intermixed in one study without acknowledging the problem type. This is problematic. Different types of WPs are presented for various student groups, in different schools or different age groups. For example, Swanson et al. (2013) investigated the role of strategy instruction and cognitive abilities on WP solving accuracy. The mathematical WPs they used were: addition, subtraction, and multiplication without any further description of the problem type. However, the available literature has already shown that different categories of WPs may lead to different solution strategies and different error types. For instance, different semantic problem types result in different errors (Vicente et al., 2007) and have a different difficulty level (LeBlanc and Weber-Russell, 1996). Obviously, different scientific studies reporting results for different student or age groups cannot be easily compared to one another when they use different WP types; it cannot be determined whether differences should be attributed to group or study manipulation or differences in the used stimulus material. In the following, we outline the major distinctions discussed in the literature. Besides the difficulty level, WPs have been categorized with regard to various other attributes. Based on standard algebra text books, Mayer (1981) categorized WPs according to their frequency. Riley et al. (1983) created four groups based on the *semantic structure* of additive arithmetic WPs (change, compare, combine, equalize) and 18 further subcategories. For instance, the change problem—where there is a start, a change, and a result state—can be subdivided into three subcategories depending on which state is the unknown.

The *mathematical content* of WPs can also serve as a basis for categorization. Algebra WPs typically require translation into a mathematical formula, whereas arithmetic WPs are solvable with simple arithmetic or even mental calculation. In contrast to arithmetic WPs, algebraic reasoning WPs share the same numerals and signs (Powell and Fuchs, 2014) and the manipulation of those numbers and signals differs based on the question or expected outcome (Kieran, 1990). However, the distinction is not that straightforward, as in some cases both methods can be applied. For instance, in a study by Van Dooren et al. (2002), future secondary school teachers preferred the use of algebra

even when an arithmetical solution seemed more evident, and some future primary school teachers rather applied arithmetical methods. Computer-aided environments have been introduced for algebraic WPs (Reusser, 1993) to support learning on "getting the formalism" and the "equation" (Nathan et al., 1992) and to allow students to generate, manipulate, and understand abstract formal expressions for WPs. However, solution approaches are not easily dissociable between arithmetic and algebraic problems. If a WP is intended to be solved with an equation, in some cases a simple arithmetic approach is enough (Gasco et al., 2014). Under some circumstances, it is even easier to solve WPs via alternative arithmetic strategies than by deriving algebraic equations. US children perform better on a story problem if it is in a money context and the numbers involve multiples of 25 (Koedinger and Nathan, 2004). While the distinction between algebra and arithmetic WPs is important for investigation and evaluation, in this review we concentrate mainly on arithmetic WPs.

*Standardized phrases* and the *idea that every problem is solvable* are other important attributes of many, but not all WPs. Textbooks generally suggest implicitly that every WP is solvable and that every numerical information is relevant (Pape, 2003). They usually provide standardized phrases and keywords that are highly correlated with correct solutions (Hinsley et al., 1977; ?). There are so-called non-standard WPs (Jimenez and Verschaffel, 2014) which can be non-solvable WPs or if they are solvable some have multiple solutions and may contain irrelevant data. In the recent literature, non-standard WPs are getting more and more attention (Yeap et al., 2005; Csikos et al., 2011). Children give a high level of incorrect answers to non-standard WPs because these seem to contradict their mathematics-related beliefs learned in the classroom. Reusser (1988) presented 97 first and second graders with the following sentence: "There are 26 sheep and 10 goats on a ship. How old is the captain?" and 76 students "solved" the problem using the numbers in the task. The rationale behind such studies is that always-solvable textbook problems with standardized phrases and including only relevant numerical information are hardly ecologically valid. Real-life WPs are not standardized, contain irrelevant information, and a solution may not always exist.

The above subcategories, which essentially characterize specific sets of WP properties, have a direct impact on human performance in WP. For space limitations, we cannot discuss the impact of all subcategories in detail, but we illustrate their impact on performance and strategies with two examples: (i) different subcategories can result in different errors, and involve different representations and processes. For example, a familiar misconception is that multiplication (Vergnaud, 2009) always makes the result larger (which is not true for  $n < 1$ ), that division makes the results smaller, and that division always involves division of the larger number by the smaller, (ii) addition problems are strongly influenced (De Corte and Verschaffel, 1987) by the semantic structure (change, compare, combine). Carpenter et al. (1981) reported that the dominant factor in determining the children's solution strategy was this semantic structure. For instance, *Change* problems [cf. the classification of Riley et al. (1983)] require the child to find the difference between the two numbers given in the problem; their nature influences the strategies

children adopt. Riley et al. (1983) illustrates this with the following examples: Change 2: “Joe had eight marbles. Then he gave five marbles to Tom. How many marbles does Joe have now?” Change 3: “Joe had three marbles. Then Tom gave him some more marbles. Now Joe has eight marbles. How many marbles did Tom give him?” Almost all the children used a subtraction strategy (e.g., counting up) to solve Change 2. For Change 3 almost all the children used an addition strategy (e.g., counting down). In sum, the subcategories introduced in this section influence both performance and the choice of solution strategies.

Indeed, solution strategies have systematically been in the focus of WP research and addressed the following questions: how do children and adults solve WPs? Why do they make different errors and at which level of the solution process they do so? Which kind of semantic representation do they create of the WP? Which skills are necessary for the solution process? The first theories on WP solution processes (Kintsch and Greeno, 1985) have drawn on the text comprehension theories of Mayer (1982) and Van Dijk and Kintsch (1983). When solving problems, the solver first integrates the textual information into an appropriate situation model or a mental representation of the situation being described in the problem, which then forms the basis for a solution strategy. This approach was further applied by (Thevenot and Oakhill, 2005; Jimenez and Verschaffel, 2014; Kingsdorf and Krawec, 2014). An important foundation of those approaches is that solving WPs is not a simple translation of problem sentences into equations (Paige and Simon, 1966). Often both WPs and the corresponding numerical problems are done without language translation (Schley and Fujita, 2014). Several researchers have focused on abstraction as a reductive process involved in the translation process in the WPs. Nathan et al. (1992) argue that WPs solving is an exercise in text processing required for understanding the problem (Cummins et al., 1988), which is highly dependent upon language comprehension skills. Successfully solving WPs has been argued to require at least three distinct processes (Nesher and Teubal, 1975): (i) understanding and constructing the relation between text and arithmetic task, (ii) linguistic understanding of the WP itself, and (iii) solving the arithmetic tasks. Typically only the latter process is assumed to be shared with common arithmetic tasks. Many students can successfully solve common arithmetic tasks and they show good text comprehension skills. Yet they fail to solve WPs correctly. This suggests that other factors like solution strategies and building up a mental model of the task also play a major role for the WP performance.

## Linguistic Complexity and Linguistic Studies

In linguistics, the notion of complexity is discussed under a range of perspectives, with particularly fruitful definitions grounded in research on language evolution (Nichols, 1990) and language acquisition (Bulté and Housen, 2012). Following the latter, it is useful to delineate linguistic complexity from propositional complexity (the amount of meaning to be expressed) and discourse-interactional complexity (the interaction of

participants in discourse). This makes it possible to zoom in on linguistic complexity as the degree to which a text at hand is elaborated and varied (Ellis, 2003, p. 340). Linguistic complexity can be analyzed with respect to all aspects of the linguistic system: from the words and their lexical and morphological aspects, via the way these words can be combined in syntax to form sentences, to the text structure, and overall discourse. Languages differ with respect to where in the linguistic system complexification is supported. For example, English makes use of word order to encode grammatical functions, whereas agglutinative languages such as Hungarian or Turkish make use of a rich morphological inventory for this and other uses. The implication of linguistic encoding differences is twofold: first, the difficulty of WPs is language-specific, thus linguistic manipulation leading to increased WP complexity in one language may not have an effect in another, more complex language. Second, the performance of language learners on WPs presented in a foreign language may be affected by the differences between the learner’s mother tongue and the language of the problem presentation. In the following two sections, we briefly summarize the main findings on aspects of linguistic complexity that affect performance.

## Structural Factors

Studies on the relation between linguistic structure and student performance on WPs have considered complexity at the micro-level of word and sentence forms as well as at the macro-level of the discourse structure of the WP passage. Early approaches addressed structural complexity in terms of basic quantitative properties of the WP text, such as the number of letters, words, sentences, mean word, and sentence length, or the proportion of complex (long) words (Searle et al., 1974; Nesher, 1976; Lepik, 1990). More linguistically motivated variables have been investigated in the context of comprehension difficulties in WPs for language learners, for the most part learners of English. At the vocabulary level, comprehension difficulties which result in problem solving difficulties for English language learners may stem from the presence of unfamiliar (low-frequency) words, polysemous words, idiomatic or culturally specific lexical references. At the sentence structure level, factors that have been shown to play a role include noun phrase length, the number of prepositional phrases and participial modifiers, the presence of passive voice and complex clause structure such as relative, subordinate, complement, adverbial, or conditional clauses (Spanos et al., 1988; Abedi et al., 1997; Abedi and Lord, 2001; Shaftel et al., 2006; Thevenot et al., 2007; Martiniello, 2008).

At the discourse structure level, specifically in terms of discourse ordering, the correspondence between the order in which numerical data is presented in the WP and the order in which it can be used to solve it has been shown to be a major predictive variable. Order-consistent problems result in better performance (Searle et al., 1974). Better performance has also been observed for simpler question wording or placing the question before the text results (Cummins et al., 1988).

## Semantic Factors

A single factor that is straightforwardly related to WP difficulty and that has been widely investigated is the presence or absence

of explicit verbal cues whose semantics hint at the expected operation and thus directly lead toward the solution. Verbal cues include words and phrases of different categories: conjunctions (“and” for addition), adverbs (“left,” “more than,” “less than” for subtraction), or determiners (“each” for multiplication). Eye tracking studies have shown that subjects tend to focus on linguistic verbal cues and perform translation directly to the mathematical operation (e.g., Hegarty et al., 1992; van der Schoot et al., 2009).

Because verbal cues so often lead to default mathematical interpretation (Nesher, 1976), even small differences in phrasing in cue words can cause significant changes in performance (LeBlanc and Weber-Russell, 1996). This is especially relevant for young children (Lean et al., 1990), who in the course of development connect words such as “join,” “add,” “get,” “find,” or “take away” with concepts such as *putting together*, *separating*, *giving away*, or *losing*. A problem can thus be reworded by adding verbal clues which make the semantic relations more salient so that the underlying mathematical relation is more explicit. For example, the WP “There are five marbles. Two of them belong to Mary. How many belong to John?” can be reworded as “There are five marbles. Two of them belong to Mary. The rest belong to John. How many belong to John?” [from Cummins (1991)]. This kind of conceptual rewording has been shown to be useful to improve children’s performance on WPs (Vicente et al., 2007). Thus changes in wording can influence representation (De Corte et al., 1985).

Semantic or object relations between the objects described in the problem also relate to difficulty. Division problems usually involve functionally related objects (e.g., *tulips–vases*) and rarely categorically related objects (e.g., *tulips–daisies*; ?). By contrast, addition for the most part involves categorically related objects. The correlation between object relations and mathematical operations has been argued to reflect a structural correspondence between semantic and mathematical relations (Bassok et al., 1998). For this reason, the semantic structure properties of a WP have been emphasized as a more important factor contributing to difficulty than the syntactic structure (Yeap and Kaur, 2001; ?). Interestingly, an effect related to information load has been observed; the presence of content irrelevant to the core solution, i.e., the presence of numerical or linguistic distractors, results in higher error rates (Muth, 1992). De Corte and Verschaffel (1987) found that the semantic structure of WPs influences children’s choice of mathematical solution strategy. In terms of the broader task context, the required or expected way of responding to the WP has a big influence, especially for the domain of multiplication and division with rational numbers as argued in De Corte et al. (1988); for example, whether students are expected to answer the problem numerically or if they only have to indicate the required operation, or whether they respond in an open way or with multiple choice.

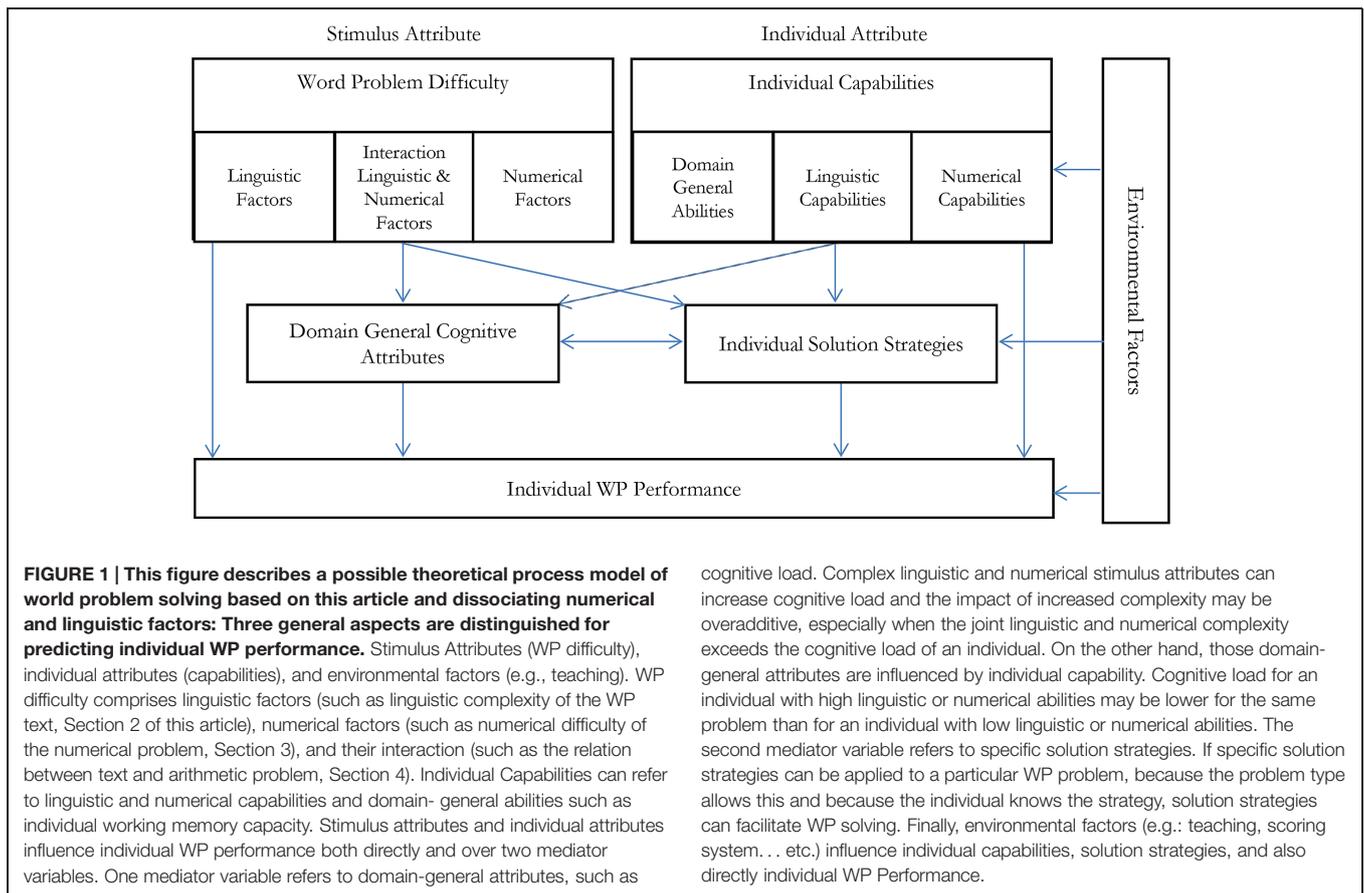
## Numerical Complexity and Numerical Studies

Arithmetic WPs have to be usually transformed mentally into an arithmetic problem and usually require an arithmetic solution

(?). This means transforming word and numbers into the appropriate operation (Neef et al., 2003). Since the arithmetic problem has to be solved in the end, numerical representations and arithmetic processes will also play an important role in the solution process. In numerical cognition, different models and representations have been proposed (e.g., Dehaene and Cohen, 1995; Nuerk et al., 2011). However, the problem here is that the literature on WP often seems (with some exceptions) to be largely in a parallel research universe to the literature on numerical cognition and arithmetic processes, so that standard models of numerical cognition are hard to apply on the existing literature. What is more, WP research on numerical factors is also affected by the scoring criteria; in some studies on WP solving, computational errors are neglected, because in many studies researchers consider a solution as correct as long as the solver has chosen the correct mathematical model (Verschaffel and De Corte, 1990). This is not the case in behavioral numerical cognition research, where the correct result is usually essential and RTs, accuracies, error types, and solution types are analyzed based on the arithmetic problem and result.

Numerical complexity can influence WP performance via at least three routes (see **Figure 1**):

- (1) Direct route: WPs with more complex arithmetic structure are more difficult independent of linguistic complexity.
- (2) Cognitive load: more complex arithmetic problems involve a higher cognitive load. For instance, carry problems are supposed to require more working memory resources. If the linguistic properties are also complex and the built-up of a mental model also requires more working memory resources, high arithmetic, and linguistic complexities could lead to over additive difficulties which could neither be explained by main effects of linguistic or numerical difficulty.
- (3) Solution strategies: multi-digit numbers are harder to process than single-digit numbers (Nuerk et al., 2011; for a review) and arithmetic complexity usually increases with numbers of digits. Thevenot and Oakhill (2005) compared the influence of processing three-digit numbers and two-digit numbers on WP solution strategies. They showed that processing numerically more complex three-digit numbers facilitated alternative strategies by the participants. The authors suggested that higher work load and working memory led to this facilitation. For our review and the model in the revised manuscript, the important point is that they resort to less effortful strategies. Similar results were observed by Brissiaud and Sander (2010), who manipulated the size and order of the numbers and operation thus resulting in two different solution strategies: (i) situation-strategy WPs that are easy to solve with informal strategies, e.g., double-counting, derived number fact, or trial-error strategy, (ii) mental arithmetic-strategy WPs are “easy to solve with mental calculation, but only when the relevant arithmetic knowledge is used.” Number magnitude and order determined which strategy was used most likely. In this review, we suggest that resorting to alternative easier strategies is not restricted to number magnitude, but could be used with any numerical variable that allows simpler solutions. For instance, if a number



bisection task were used in a text problem, we would also suggest that participants resort to easier strategies (e.g., checking the parities of the outer number), when the bisection problem gets more complex (e.g., larger interval, decade crossing etc.).

Nevertheless, some distinctions of numerical processes can be made in our review of the WP literature and are therefore proposed as an initial step in this review. Note that in our view this is not the end of the integration of numerical cognitive research and WP research, but rather just a beginning. For an overview of the investigation of specific numerical processes in current WP research, we suggest categorizing them into five categories:

- (i) the property of numbers (parity, single digit/multi digit, problem size, ties, type of number, role of the number, number magnitude),
- (ii) required operation (type, number)
- (iii) mathematical solution strategies (larger number, place, automatic fact retrieval, position of the unknown),
- (iv) relevance of the information.
- (v) other numerical processes and representations

## Number Properties

While some studies have shown an effect of numerical complexity, from a numerical cognition view it is surprising that

actually the arithmetic complexity has rarely been systematically considered as an isolated factor in WPs, although it is frequently examined in other arithmetic problems or simply the description of numbers is missing, e.g., De Corte et al. (1990). For instance, *parity* attributes are rarely considered in WPs, although in children it influences task performance and strategy choice in arithmetic tasks. For instance, in the number bisection task (Is the middle number  $Y$  the exact mean of  $X$  and  $Z$  in  $X\_Y\_Z?$ ), parity influences performance. Trials with unequal parities of  $X$  and  $Z$  are easier to solve than trials with equal parities (Nuerk et al., 2002). We suggested that this is due to a change in strategy. In trials with unequal parity (e.g., 21\_25\_28), it is impossible that the middle number is the mean, because the mean of numbers with unequal parity is not an integer number (and only integers were used in the experiment). Therefore, participants may change their strategy after they discovered unequal parities and may not compute further to find out whether the middle number is really the mean. A later fMRI study (Wood et al., 2008) corroborated this assumption. In the easier unequal parity (“impossible”) condition, we observed more activation in the right ventrolateral prefrontal cortex, which is activated in cognitive set changes or when participants generate alternative solutions for a task. Thus, parity can influence performance and solution strategies in arithmetic. This seems not only the case in the bisection task, which is to our knowledge rarely used in WP research, but also in standard operations like addition and subtraction. A review by

Hines (2013) suggests that parity influences the difficulty of addition and subtraction, but not multiplication, and tasks containing odd numbers are more difficult than with even ones. Such parity effects have received little attention in WP research so far. Furthermore, it seems that most WPs, especially for children, contain *single-digit numbers*; e.g., each answer was in the range of 1–9, e.g., in Lean et al. (1990), or Powell and Fuchs (2014), only a few use *multi-digit numbers* (Haghverdi et al., 2012). In Nesher (1976) the range of numbers is smaller than 100, contained division two-digit numbers into one-digit number.

Explanations why the studies have chosen specific numbers, e.g., mentioning problem size, are rare. De Corte et al. (1990) and Orrantia et al. (2010) controlled for the number of sentences; the size of the numbers given in the problems. In the study of van der Schoot et al. (2009) the final answers were between 14 and 40, included no fraction, no negative number, no numerical value twice, and none of the possible answers resulted in another. However, different types of numbers were presented in WPs in some more studies: (i) fraction (Raduan, 2010), (ii) whole number, (iii) decimal number (Haghverdi et al., 2012); and their effect has been rarely investigated. Koedinger and Nathan (2004) found an effect for decimal numbers: “however we also observed a smaller situation facilitation effect whereby story performance was better than word equation performance under certain conditions: namely dealing with decimal numbers.”

The mixed use of single- and multi-digit numbers is problematic because in the last 15 years, numerous numerical cognition studies have shown that single-digit number processing cannot easily be generalized to multi-digit number processing, e.g., Nuerk et al. (2001); for reviews see Nuerk and Willmes (2005) and Nuerk et al. (2015). Nuerk et al. (2015) have identified 17 numerical effects linked to different numerical representation, which are specific for multi-digit number processing and which cannot be explained by single-digit number representations. Also even the same effects are different for single- and multi-digit numbers. For instance, Ashkenazi et al. (2009) have shown that the distance effect for two-digit numbers differentiates between dyscalculic and typically developing children. The sometimes seemingly arbitrary mix of single-digit and multi-digit number use in WP research is therefore not reasonable in our view given the state of numerical cognition research and the major differences between processing those different number types. The *role of a number* within an operation also influences WP complexity (De Corte et al., 1988). For example, in the case of addition the role means: addend, minuend or by multiplication: multiplicand, multiplier. One important finding from recent research on multiplication WPs is that children’s performances are strongly affected by the nature of the multiplier whether, e.g., it is an integer, decimal larger than 1 or a decimal smaller than 1. On the other hand, the size of the multiplicand has little or no effect on problem difficulty. De Corte et al. (1988) stated that “two multiplication problems with the same mathematical, semantic, and surface structure but different in terms of the nature of the given numbers can elicit very distinct levels of problems difficulty.” Indeed, this corresponds to recent findings that relatedness and consistency heavily influence the ease with which a multiplication

problem can be solved cf. for relatedness (Domahs et al., 2006, 2007) and for consistency Verguts and Fias (2005).

Despite the major role of number properties in numerical cognition, number property has not been investigated extensively in the WPs (Fuchs et al., 2009). Nevertheless, numbers seem to play a major role. For instance, De Corte and Verschaffel (1986) observed that in their eye tracking study there was a relatively strong focus on the numbers in the problem. Twenty-five percent of the total solution time was spent in the two small number areas. However, major number properties of numerical cognitions research such as number magnitude are rarely systematically considered in WP research. In our view, more dialog between fields, – numerical cognition and WP research – seems necessary.

### Required Operation

Carrying out operations are necessary steps in solving arithmetic WPs. Operations have been used extensively in WPs. Most errors seem to originate from people’s failure to understand the language of WPs, i.e., the linguistic embedding of the calculation problem (Schumacher and Fuchs, 2012), and arithmetic computation errors themselves (Raduan, 2010; Kingsdorf and Krawec, 2014). Some errors may result from correct calculation performed on incorrect problem representation (Lewis and Mayer, 1987) and different operations may lead to different solution strategies. The most usual operation used in WP experiments are addition and subtraction (Carpenter et al., 1984; De Corte et al., 1988; Schumacher and Fuchs, 2012). Even the classification of Riley et al. (1983) was made for elementary addition and subtraction. Research in the 1980s and 1990s concentrated on how children learn to do one step addition and subtraction problems involving small whole numbers; see the review from Vicente et al. (2007). Later, the focus was more on the multiplication WPs or mixed WPs – e.g., Swanson (2004). Greer (1992) presented a framework categorization of multiplication and division WPs on the basis of the types of quantities involved (positive integers, fraction, and decimals) as models of situation. The semantic problem structure also influences the solution strategies for addition and subtraction.

Choosing the correct operation strongly depends on the type of the given numbers in the problem (De Corte et al., 1990). As already shortly outlined in above subsection on problem types, there is a huge body of research on what makes addition, subtraction, or multiplication problems difficult. Carry operations (e.g.,  $28 + 47$ ; the decade value 1 from the unit sum 15 has to be carried over to the decade sum) have long been known to make multi-digit addition more difficult in children and adults; see Nuerk et al. (2015) for a review. However, solution strategies differ between children and adults – eye movement data suggest that in a choice reaction task elementary school children always compute and search for the correct results, while adults seem to also decide based on the rejection of the incorrect result. What is more, even within the carry operations at least three different cognitive processes can be identified for adults: unit sum calculation, carry detection, and carry execution (Moeller et al., 2011). Inability to execute one of these processes may lead to worse performance in carry problems in particular. Carry addition problems seem to

require larger working memory resources (Ashcraft, 1995; Furst and Hitch, 2000). If cognitive load/working memory demand is high, because both the linguistic and the numerical complexity of the WP are large, this may lead to over additive problems in the domain-general processing stages involved in WP solving — see **Figure 1**, for an elaboration. For multiplication, we know that relatedness, ties, whether a problem stems from the 0, 1, 2, 5, or 10 row (Josta et al., 2009), or consistency influence the difficulty of a multiplication problem (Domahs et al., 2006, 2007). Although such factors have been extensively studied in numerical cognition research, they are – to the best of our knowledge – rarely considered in WP research. Since we know that these factors make the arithmetic computation, which is part of the WP solution, this lack of consideration is again problematic in our view.

### Mathematical Solution Strategies

Mathematical solution strategy variations have been studied extensively, and can be a function of linguistic factors like wording, semantic categories and propositions. However, how individuals come up with mathematical solution strategies can be also be influenced by numerical factors like number magnitude (Thevenot and Oakhill, 2005). Such variables, which are independent of other factors, make WPs harder and/or influence numerical representations, have rarely been studied. The position/place of the unknown variable has an effect on representation (Garcia et al., 2006). Even studies about working memory also investigated the position of the unknown variable (Swanson, 2004). The strategy of counting on from larger is easier if the bigger number is represented first (Wilkins et al., 2001). Even for adults:  $4 + 2 = 6$ , and  $2 + 4 = 6$ , which are mathematically equivalent, may psychologically imply different meanings (Kaput, 1979). The sequence of the numbers, e.g., whether a problem starts with the smaller or with the larger number (Verschaffel and De Corte, 1990), the position of the numbers and particular words (Schumacher and Fuchs, 2012) influence children's solution of elementary addition and subtraction problems. For example, in change problems children typically look for a specific number to begin with, depending on task features, like the first mentioned number (Lean et al., 1990; Wilkins et al., 2001), the type of problem (start or change set), and the size of the numbers (Verschaffel and De Corte, 1990).

Arithmetic fact retrieval is a well researched ubiquitous strategy in numerical cognition but less so in the domain of WPs. Orrantia et al. (2010) found that arithmetic fact retrieval is not limited to simple addition, but also possible in other tasks, such as single-digit arithmetic WPs. Fuchs et al. (2009) investigated so called “Number combination.” This means simple arithmetic problems that can be solved via counting or decomposition strategies or committed to long term memory for automatic retrieval. Here, arithmetic fact retrieval had to be differentiated from other strategies on three levels: operational, items difficulty, and individual differences. These numerical factors influence solution strategies in arithmetic and WPs as well. Decomposition and counting require more working memory and therefore leave less resources for the built-up and maintenance of a text situations model. However, both individual and stimulus differences should also be considered. For instance, Grabner et al. (2009)

showed in an fMRI study that not only problem but also individual strategy choice contributed to fact retrieval processes when solving multiplications.

### Information Relevance and Step-Wise Problem Processing

One relatively extensively studied factor in WPs is the relevance of the information. Individuals have to extract the relevant information from the text in order to carry out the correct solution. Secondary information distracts people from recognizing the underlying mathematical relations (Schley and Fujita, 2014). This extra information may also be presented in the form of an extra number or an extra operational step – one-step (i.e., one calculation step has to be performed) and two-step problems (i.e., two calculation steps have to be performed). Problem complexity increases with the addition of steps (Terao et al., 2004), as well as the addition of irrelevant information to the problem (Kingsdorf and Krawec, 2014) Presence of extraneous information and the need for an extra step reduced the accuracy of the students' solutions, because students believe that all of the numbers in a WP should be used. All other factors being kept constant, two-step problems are much more error-prone than one-step problems (Muth, 1992). However, it cannot be concluded that the reason for two-step problems being more difficult is arithmetic complexity, because in two-step problems, the WP has also become more difficult linguistically as it usually contains more phrases and semantic distractors.

### Other Numerical Processes and Representation

Several other numerical processes and representations have not been investigated in WPs. For instance, as shortly outlined above, one major factor in simple calculation problems, which can be studied in isolation, is the presence or absence of a carry operation. Children and adults take longer and commit more errors when computing the solution to a sum for which adding the units leads to a change in the number of 10s (e.g.,  $14 + 9 = 23$ ; Furst and Hitch, 2000; Deschuyteneer et al., 2005) than when it does not (e.g.,  $11 + 12 = 23$ ). This effect is known as the carry effect; in carry problems, a one needs to be carried from the unit slot to the decade slot. The carry effect is influenced by various processes, but even by language structure (Goebel et al., 2014). Language influences on the difficulty of the numerical computations within a WP have to our knowledge not been studied. Other central topics of numerical cognition such as, e.g., number and symbol sense contribute to WP solving are also open questions (MacGregor and Price, 1999). We have chosen some selected variables/factors, which have been investigated in the WP research.

### Connecting Linguistic and Mathematical Factors

There are so many linguistic influences on numerical cognition and arithmetic that this justifies a special issue like this. For instance, number word structure seems to play an essential

role. Children growing up with regular number word structure usually perform better in variety of numerical tasks from basic verbal counting up to arithmetic, e.g., Miller et al. (1995) or Dowker et al. (2008). In addition, the consistency of the order of the number word system and the Arabic number influences transcoding (Zuber et al., 2009; Pixner et al., 2011a; Imbo et al., 2014) number comparison (Nuerk et al., 2005; Pixner et al., 2011b; Klein et al., 2013; Moeller et al., 2014) calculation (Goebel et al., 2014); see also (Brybaert et al., 1998; Colomé et al., 2010). In addition, reading direction influences numerical processes like the SNARC effect (Shaki et al., 2009; Fischer and Shaki, 2014); see Goebel et al. (2011) for reviews. Finally, grammatical and syntactic properties of elementary number words influence early number acquisition (Sarnecka, 2013) and spatial-numerical representations (Roettger and Domahs, 2015). The linguistic influence on numerical cognition is hardly debatable any more. In fact, Lachmair et al. (2014) argue for a connection of language and words, O'Neill (2013) states that the link between language and mathematics might originate from the same roots, and "required abilities are not that split up as we think," and MacGregor and Price (1999) also argue that between language and mathematics in WPs there is deep connection: "that the cognitive ability that drives symbol processing is the connection between language and maths." Nevertheless, systematic variation of both linguistic and numerical factors in WPs is scarce – though Bassok et al. (1998) already found that semantic relations between objects in the text of mathematical WPs were highly positively correlated with arithmetic operations that took these objects as arguments. Neural correlates of visualization and verbalization during arithmetic WP study also suggest that mental arithmetic in WPs is influenced by language processing (Zarnhofer et al., 2013).

Word problems require some connection between linguistic and mathematical understanding by the very nature of the task, because at least children do not have a repertoire of "highly automatized schemata" for representing the different problem types (Garcia et al., 2006). Therefore, it is not surprising that children make more errors when solving WPs compared to number problems (Geary, 1996; Koedinger and Nathan, 2004). Children are able to solve several types of addition and subtraction problems before they start formal schooling (De Corte and Verschaffel, 1987; Lean et al., 1990), and understand numerical concepts before seeing WPs in their curricula (Garcia et al., 2006). Therefore, most studies implicitly assume that problem solvers always have the necessary basic arithmetic skills, even in the case of children. This may lead to the misconception that numbers may play a lesser role than they actually do and factors other than computational skills are a major source of difficulty with WPs (Nesher, 1976; Reusser, 1993). In this aspect, it is also important to note that difficulties in solving WPs have been reported that could be neither attributed to the lack of general reading comprehension skills nor to the lack of general mathematical skills (Hegarty et al., 1995). Nevertheless, linguistics and numerical factors are usually not independently manipulated in WPs and not even dissociated by other means (e.g., regressions). What is more, their interaction is rarely studied [for an exception, see Verschaffel and De Corte (1990)].

## Lexical Consistency Effect

One of the few frequently studied factors examining the relation between text and arithmetic problems is lexical inconsistency. Some WPs contain linguistic markers as "less" or "more." In the direct translation strategy (Hegarty et al., 1995) students simply associate "less" with subtraction and "more" with addition. They search for linguistic markers and keywords. In the problem model strategy, they construct a mental model of the problem and plan their solution on the basis of this model. Successful learners are more likely to employ the problem model strategy; they focus more on variables names and relational terms and successful problem solvers re-read the text less frequently (Pape, 2003) in the eye-tracking studies. Unsuccessful learners, on the other hand, seem to rely on the direct translation strategy; they focus on numerals and on relational terms, and linguistics markedness in the (Hegarty et al., 1992) eyetracking study. This leads to wrong solutions in lexically inconsistent texts, where "more" is associated with subtraction and "less" with addition. To give an example for lexical inconsistency, consider the following WP adapted from Boonen et al. (2013) "At the grocery store, a bottle of olive oil costs 7 €. That is 2 € more than at the supermarket. How much will [a bottle of olive oil] cost in the supermarket?" The anticipated difficulty in comprehension and finding the correct solution is due to the fact that the adverb "more" evokes the concept of addition, but the correct solution is not  $7 + 2$  but  $7 - 2$ , given the way the text is organized. Verschaffel et al. (1992) found such a reaction time consistency effect for children but not for adults. Nesher (1976) and Lean et al. (1990) obtained similar results in experiments with groups of non-disadvantaged children and students, showing that linguistic semantic consistency with respect to the required mathematical operation is an important determinant of task difficulty. Inconsistent language results in a high error rate and longer response time (Hegarty et al., 1992), even in Verschaffel (1994) retelling one-step compared WPs showed a strong evidence for the consistency hypothesis. Students made ~13% more reversal errors on inconsistent than on consistent language problems and the difficulty of comprehending inconsistent-language problems were increased when the correct arithmetic operation was an increase. However, the literature is inconsistent if the consistency effect is present in both students and children. Children find it easier to convert the relation term "more than" into subtraction operation than the relational term "less than" into an addition operation (Lewis and Mayer, 1987; Verschaffel et al., 1992; Pape, 2003; van der Schoot et al., 2009).

When neither reading comprehension nor arithmetic skills alone can explain failure to solve WPs, a possible explanation is that linguistic complexity and numerical complexity rely on the same resources (e.g., working memory). The premise is that there is not an absolute atomic concept of difficulty for WPs. Rather; there are multiple linguistic and numerical factors which contribute to a problem's complexity. It is a combination of these factors that might make a problem additively more or less difficult because they exert demands on more general resources like working memory. Generally, problem solving performance is related to the ability of reducing the accessibility of no target and irrelevant information in the memory

(Passolunghi and Siegel, 2001). Working memory contributes to early arithmetic performance, and studies also show that this extends to WP solving (Lee et al., 2004) due to semantic memory representation “less than” which is more complex than “more than.” Changes in the structure of the text has more demand on the working. It has been suggested that WPs in general are related to working memory (Swanson et al., 1993). This will probably also be influenced by instruction specifying how participants have to solve a WP, and the method of evaluation, and scoring system. In Van Dijk and Kintsch’s (1983) model of reading comprehension, working memory is used to keep a number of text propositions active simultaneously. In particular, working memory has been related to each single component mentioned above, such as text-problem relation, the linguistic complexity, and the arithmetic complexity.

## Future Direction, Open Questions

Word problem difficulty is influenced by the complexity of linguistic factors, numerical factors, and their interrelation. To better understand the difficulty of WPs, it would be desirable to manipulate such variables and their interaction following the principle of isolated variation. To support a systematic investigation, the variables to be manipulated also need to be discussed against the backdrop of the relevant conceptual and empirical issues in the underlying fields, linguistics, and numerical cognition. This has too rarely been the case in the past. For instance, in the earlier studies on algebra WPs, the linguistic cues are of mixed categories (adverbs, verbs, nouns, etc.) and the effect of the complexity of syntactic structures is not taken into account. Similarly, numerical complexity like basic number properties (e.g., magnitude, place-value processing for multi-digit numbers) or the complexity of underlying arithmetic computations (e.g., carry effects for addition, relatedness, or consistency effects for multiplication) are often neglected. WP research would be well advised to take into account the foundational categories, properties and findings of both numerical cognition and linguistics when it examines which WPs are difficult for which groups and why. Not only the main effects of numerical and linguistic complexity should be studied, but also their interaction. To make

the relevant aspects explicit, **Figure 1** sketches an overall process model of WP solving.

The joint investigation of linguistic and numerical processes also needs to take into account joint moderator variables such as working memory in order to explore the possible interactions between them. Since working memory affects all components of complexity of a WP, the difficulties triggered may not be simply additive, but also interactive. The resolution of linguistic and numerical difficulties may rely on the same processing stages and resources (Sternberg, 1969). To investigate this, more collaboration between linguists and numerical cognition researchers would be desirable.

Finally, we suggest a differential-psychological approach to WP research. Different students may have a problem with different types of WPs. Linguistically rather weak students may have problems with linguistically complex WPs, and arithmetically rather weak students with arithmetically complex problems. Undifferentiated presentation of WPs in experiments will not provide sufficient information about which skills and processes an individual child should practice. Only with such differentiation on an item level (as regards linguistic and numerical complexity and their interrelation) and on an individual level (as regards linguistic and numerical skills and general cognitive abilities) will it be possible to understand why a particular child has its individual difficulties with particular WP types. Such an understanding, however, is essential to promote tailored learning of one of the most difficult arithmetic problem types that students encounter in school.

## Acknowledgments

This research was funded by the LEAD Graduate School [GSC1028], a project of the Excellence Initiative of the German federal and state governments. MW is a Junior Research Group Leader of the LEAD Graduate School and her work is supported by the Institutional Strategy of the University of Tübingen (Deutsche Forschungsgemeinschaft, ZUK 63). GD is a doctoral student of the LEAD Graduate School. We acknowledge support by the Deutsche Forschungsgemeinschaft and the Open Access Publishing Fund of the University of Tübingen.

## References

- Abedi, J., and Lord, C. (2001). The language factor in mathematics tests. *Appl. Meas. Educ.* 14, 219–234. doi: 10.1207/S15324818AME1403\_2
- Abedi, J., Lord, C., and Plummer, J. (1997). *Final Report of Language Background as a Variable in NAEP Mathematics Performance*. Tech. Rep. No. 429. Los Angeles, CA: University of California, Center for the Study of Evaluation/National Center for Research on Evaluation, Standards, and Student Testing.
- Ashcraft, M. H. (1995). Cognitive psychology and simple arithmetic: a review and summary of new directions. *Math. Cogn.* 1, 3–34.
- Ashkenazi, S., Rubinsten, O., and Henik, A. (2009). Attention, automaticity, and developmental dyscalculia. *Neuropsychology* 23, 535–540. doi: 10.1037/a0015347
- Bassok, M., Chase, V. M., and Martin, S. A. (1998). Adding apples and oranges: alignment of semantic and formal knowledge. *Cogn. Psychol.* 35, 99–134. doi: 10.1006/cogp.1998.0675
- Berends, I. E., and van Lieshout, E. C. D. M. (2009). The effect of illustrations in arithmetic problem-solving: effects of increased cognitive load. *Learn. Instr.* 19, 345–353. doi: 10.1016/j.learninstruc.2008.06.012
- Boonen, A. J. H., Van Der Schoot, M., Van Wesel, F., De Vries, M. H., and Jolles, J. (2013). What underlies successful word problem solving? A path analysis in sixth grade students. *Contemp. Educ. Psychol.* 38, 271–279. doi: 10.1016/j.cedpsych.2013.05.001
- Boonen, A. J. H., Van Wesel, F., Jolles, J., and Van Der Schoot, M. (2014). The role of visual representation type, spatial ability, and reading comprehension in word problem solving: an item-level analysis in elementary school children. *Int. J. Educ. Res.* 68, 15–26. doi: 10.1016/j.ijer.2014.08.0010883-0355
- Brissiaud, R., and Sander, E. (2010). Arithmetic word problem solving: a situation strategy first framework. *Dev. Sci.* 13, 92–107. doi: 10.1111/j.1467-7687.2009.00866.x
- Brybaert, M., Fias, W., and Noël, M. P. (1998). The Whorfian hypothesis and numerical cognition: is “twenty-four” processed in the same way as “four-and-twenty”? *Cognition* 66, 51–77. doi: 10.1016/S0010-0277(98)00006-7

- Bulté, B., and Housen, A. (2012). "Defining and operationalising L2 complexity," in *Dimensions of L2 Performance and Proficiency – Investigating Complexity, Accuracy and Fluency in SLA*, eds A. Housen, F. Kuiken, and I. Vedder. (Amsterdam: John Benjamins), 21–46. doi: 10.1075/llt.32.02bul
- Carpenter, T. P., Hieben, J., and Moser, J. (1981). The effect of problem structure on first-grader's initial solution processes for simple addition and subtraction problems. *J. Res. Math. Educ.* 2, 27–39. doi: 10.2307/748656
- Carpenter, T. P., Moser, T. P., and James, M. (1984). The acquisition of addition and subtraction concepts in grades one through three. *J. Res. Math. Educ.* 15, 172–202. doi: 10.2307/748348
- Colomé, A., Laka, I., and Sebastiaen-Galles, N. (2010). Language effects in addition: how you say it counts. *Q. J. Exp. Psychol.* 63, 965–983. doi: 10.1080/17470210903134377
- Csikos, C., Kelemen, R., and Verschaffel, L. (2011). Fifth-grade students' approaches to and beliefs of mathematics word problem solving: a large sample Hungarian study. *ZDM* 43, 561–571. doi: 10.1007/s11858-011-0308-7
- Cummins, D. D. (1991). Children's interpretations of arithmetic word problems. *Cogn. Instr.* 8, 261–289. doi: 10.1207/s1532690xci0803\_2
- Cummins, D. D., Kintsch, W., Reusser, K., and Weimer, R. (1988). The role of understanding in solving word problems. *Cogn. Psychol.* 20, 405–438. doi: 10.1016/0010-0285(88)90011-4
- De Corte, E., and Verschaffel, L. (1986). Eye-movement data as access to solution processes of elementary addition and subtraction problems. *Paper Presented at the Annual Meeting of the American Educational Research Association*, San Francisco.
- De Corte, E., and Verschaffel, L. (1987). The effect of semantic structure on first graders' strategies for solving addition and subtraction word problems. *J. Res. Math. Educ.* 18, 363–381. doi: 10.2307/749085
- De Corte, E., Verschaffel, L., and De Win, L. (1985). Influence of rewording verbal problems on children's problem representations and solutions. *J. Educ. Psychol.* 77, 460–470. doi: 10.1037/0022-0663.77.4.460
- De Corte, E., Verschaffel, L., and Pauwels, A. (1990). Influence of the semantic structure of word problems on second graders eye movements. *J. Educ. Psychol.* 82, 359–365. doi: 10.1037/0022-0663.82.2.359
- De Corte, E., Verschaffel, L., and Van Coillie, V. (1988). Influence of number size, problem structure, and response mode on children's solutions of multiplication word problems. *J. Math. Behav.* 7, 197–216.
- Dehaene, S., and Cohen, L. (1995). Towards an anatomical and functional model of number processing. *Math. Cogn.* 1, 83–120.
- Dellarosa, D. (1986). A computer simulation of children's arithmetic word-problem solving. *Behav. Res. Methods Instrum. Comput.* 18, 147–154. doi: 10.3758/BF03201014
- Deschuyteneer, M., De Rammelaere, S., and Fias, W. (2005). The addition of two-digit numbers: exploring carry versus no-carry problems. *Psychol. Sci.* 47, 74–83.
- Domahs, F., Delazer, M., and Nuerk, H.-C. (2006). What makes multiplication facts difficult: problem size or neighbourhood consistency? *Exp. Psychol.* 53, 275–282. doi: 10.1027/1618-3169.53.4.275
- Domahs, F., Janssen, U., Schlesewsky, M., Ratinckx, E., Verguts, T., Willmes, K., et al. (2007). Neighborhood consistency in mental arithmetic: behavioral and ERP evidence. *Behav. Brain Funct.* 3, 66. doi: 10.1186/1744-9081-3-66
- Dowker, A., Bala, S., and Lloyd, D. (2008). Linguistic influences on mathematical development: how important is the transparency of the counting system. *Philos. Psychol.* 21, 523–538. doi: 10.1080/09515080802285511
- Ellis, R. (2003). *Task-based Language Learning and Teaching*. Oxford: Oxford University Press.
- Fischer, M. H., and Shaki, S. (2014). Spatial associations in numerical cognition – from single digits to arithmetic. *Q. J. Exp. Psychol.* 67, 1461–1483. doi: 10.1080/17470218.2014.927515
- Fite, G. (2002). Reading and math: what is the connection? A short review of the literature. *Kansas Science Teacher* 14, 7–11.
- Fuchs, L. S., Powell, S. R., Seethaler, P. M., Cirino, P. T., Fletcher, J. M., Fuchs, D., et al. (2009). Remediating number combination and word problem deficits among students with mathematics difficulties: a randomized control trial. *J. Educ. Psychol.* 101, 561–576. doi: 10.1037/a0014701
- Furst, A., and Hitch, G. J. (2000). Separate roles for executive and phonological components of working memory in mental arithmetic. *Mem. Cogn.* 28, 774–782. doi: 10.3758/BF03198412
- Garcia, A., Jimenez, L., and Hess, S. (2006). Solving arithmetic word problems: an analysis of classification as a function of difficulty in children with and without arithmetic LD. *J. Learn. Disabil.* 39, 270–281. doi: 10.1177/00222194060390030601
- Gasco, J., and Villarroel, J. D. (2014). The motivation of secondary school students in mathematical word problem solving. *Electron. J. Res. Educ. Psychol.* 12, 83–106. doi: 10.14204/ejrep.32.13076
- Gasco, J., Villarroel, J. D., and Zuazagoitia, D. (2014). Different Procedures for Solving Mathematical Word Problems in High School. *Int. Educ. Stud.* 7, 77. doi: 10.5539/ies.v7n7p77
- Geary, D. C. (1996). *Children's Mathematical Development: Research and Practical Applications*. Washington, DC: American Psychological Association.
- Goebel, S., Moeller, K., Pixner, S., Kaufmann, L., and Nuerk, H.-C. (2014). Language affects symbolic arithmetic in children: the case of number word inversion. *J. Exp. Child Psychol.* 119, 17–25. doi: 10.1016/j.jecp.2013.10.001
- Goebel, S. M., Shaki, S., and Fischer, M. H. (2011). The cultural number line: a review of cultural and linguistic influences on the development of number processing. *J. Cross Cult. Psychol.* 42, 543–565. doi: 10.1177/0022022111406251
- Grabner, R. H., Ansari, D., Koschutnig, K., Reishofer, G., Ebner, F., and Neuper, C. (2009). To retrieve or to calculate? Left angular gyrus mediates the retrieval of arithmetic facts during problem solving. *Neuropsychologia* 47, 604–608. doi: 10.1016/j.neuropsychologia.2008.10.013
- Greer, B. (1992). "Multiplication and division as models of situations," in *HandBook of Research on Mathematics Teaching and Learning*, ed. D. A. Grouws (New York, NY: Macmillan Publishing Company), 276–295.
- Haghverdi, M., Semnani, A. S., and Seifi, M. (2012). The relationship between different kinds of students' errors and the knowledge required to solve mathematics word problems. *Bolema Bol. Educ. Matem.* 26, 649–666. doi: 10.1590/S0103-636X2012000200012
- Hegarty, M., Mayer, R. E., and Green, C. E. (1992). Comprehension of arithmetic word problems: evidence from students' eye fixations. *J. Educ. Psychol.* 84, 76–84. doi: 10.1037/0022-0663.84.1.76
- Hegarty, M., Mayer, R. E., and Monk, C. A. (1995). Comprehension of arithmetic word problems: a comparison of successful and unsuccessful problem solvers. *J. Educ. Psychol.* 87, 18–32. doi: 10.1037/0022-0663.87.1.18
- Hines, T. M. (2013). Parity influences the difficulty of simple addition and subtraction but not multiplication problems in children. *Psychol. Rep.* 113, 1048–1065. doi: 10.2466/10.11.PR0.113x16z4
- Hinsley, D. A., Hayes, J. R., and Simon, H. A. (1977). "From words to equations: meaning and representation in algebra word problems," in *Cognition Processes in Comprehension*, ed. P. S. Carpenter (Hillsdale, NJ: Erlbaum), 89–100.
- Imbo, I., Bulcke, C. V., De Brauwer, J., and Fias, W. (2014). Sixty-four or four-and-sixty? The influence of language and working memory on children's number transcoding. *Front. Psychol.* 5:313. doi: 10.3389/fpsyg.2014.00313
- Jimenez, L., and Verschaffel, L. (2014). Development of children's solutions of non-standard arithmetic word problem solving non-standard arithmetic word problems. *Rev. Psicodidactica* 19, 93–123. doi: 10.1387/RevPsicodidact.7865
- Josta, K., Khadera, P., Burkea, M., Bienb, S., and Röslera, F. (2009). Dissociating the solution processes of small, large, and zero multiplications by means of fMRI. *Neuroimage* 46, 308–318. doi: 10.1016/j.neuroimage.2009.01.044
- Kaput, J. J. (1979). "Mathematics and learning: roots of epistemological status," in *Cognitive Process Instruction*, eds J. Clement and J. Lochhead (Philadelphia, PA: Franklin Institute Press), 289–303.
- Kieran, C. (1990). "Cognitive processes involved in learning school algebra," in *Mathematics and Cognition*, eds P. Neshier and J. Kilpatrick (New York, NY: Cambridge University Press), 96–112.
- Kingsdorf, S., and Krawec, J. (2014). Error analysis of mathematical word problem solving across students with and without learning disabilities. *Learn. Disabil. Res. Pract.* 29, 66–74. doi: 10.1111/ldrp.12029
- Kintsch, W., and Greeno, J. G. (1985). Understanding and solving word arithmetic problems. *Psychol. Rev.* 92, 109–129. doi: 10.1037/0033-295X.92.1.109
- Klein, E., Moeller, K., Kettl, D., Zauner, H., Wood, G., Willmes, K., et al. (2013). Object-based neglect in number processing. *Behav. Brain Funct.* 9, 5. doi: 10.1186/1744-9081-9-5
- Koedinger, K. R., and Nathan, M. J. (2004). The real story behind story problems: effects of representations on quantitative reasoning. *J. Learn. Sci.* 13, 129–164. doi: 10.1207/s15327809jls1302\_1

- Lachmair, M., Dudschig, C., De La Vega, L., and Kaup, B. (2014). Relating numeric cognition and language processing: do numbers and words share a common representational platform? *Acta Psychol.* 148, 107–114. doi: 10.1016/j.actpsy.2013.12.004
- Lean, G. A., Clemens, M. A., and Del Campo, G. (1990). Linguistic and pedagogical factors affecting children's understanding of arithmetic word problems: a comparative study. *Educ. Stud. Math.* 21, 165–191. doi: 10.1007/BF00304900
- LeBlanc, M. D., and Weber-Russell, S. (1996). Text integration and mathematical connections: a computer model of arithmetic word problem solving. *Cogn. Sci.* 20, 357–407. doi: 10.1207/s15516709cog2003\_2
- Lee, K., Ng, S. F., Ng, E. L., and Lim, R. Y. (2004). Working memory and literacy as predictors of performance on algebraic word problems. *J. Exp. Child Psychol.* 89, 140–158. doi: 10.1016/j.jecp.2004.07.001
- Lepik, M. (1990). Algebraic word problems: role of linguistic and structural variables. *Educ. Stud. Math.* 21, 83–90. doi: 10.1007/BF00311017
- Lewis, A. B., and Mayer, R. E. (1987). Students' miscomprehension of relational statements in arithmetic word problems. *J. Educ. Psychol.* 79, 363–371. doi: 10.1037/0022-0663.79.4.363
- MacGregor, M., and Price, E. (1999). An exploration of aspects of language proficiency and algebra learnin. *J. Res. Math. Educ.* 30, 449–467. doi: 10.2307/749709
- Martin, S. A., and Bassok, M. (2005). Effects of semantic cues on mathematical modeling: evidence from word-problem solving and equation construction tasks. *Mem. Cognit.* 33, 471–478. doi: 10.3758/BF03193064
- Martiniello, M. (2008). Language and the performance of english-language learners in math word problems. *Harv. Educ. Rev.* 78, 333–368.
- Mayer, R. E. (1981). Frequency norms and structural analysis of algebra story problems into families, categories, and templates. *Instr. Sci.* 10, 135–175. doi: 10.1007/BF00132515
- Mayer, R. E. (1982). Memory for algebra story problems. *J. Educ. Psychol.* 74, 199–216. doi: 10.1037/0022-0663.74.2.199
- Miller, K., Smith, C. M., Zhu, J., and Zhang, H. (1995). Preschool origins of cross-national differences in mathematical competence: the role of number-naming systems. *Psychol. Sci.* 6, 6–60. doi: 10.1111/j.1467-9280.1995.tb00305.x
- Moeller, K., Klein, E., Kucian, K., and Willmes, K. (2014). Numerical development - from cognitive functions to neural underpinnings. *Front. Psychol.* 5:1047. doi: 10.3389/fpsyg.2014.01047
- Moeller, K., Klein, E., and Nuerk, H.-C. (2011). (No) small adults - children's processing of carry addition problems. *Dev. Neuropsychol.* 36, 702–720. doi: 10.1080/87565641.2010.549880
- Muth, K. D. (1992). Extraneous information and extra steps in arithmetic word problems. *Contemp. Educ. Psychol.* 17, 278–285. doi: 10.1016/0361-476X(92)90066-90068
- Nathan, M. J., Kintsch, W., and Young, E. (1992). A theory of algebra-word-problem comprehension and its implications for the design of learning environments. *Cogn. Instr.* 9, 329–389. doi: 10.1207/s1532690xci0904\_2
- Neef, N. A., Nelles, D. E., Iwata, B. A., and Page, T. J. (2003). Analysis of precurent skills in solving mathematics story problems. *J. Appl. Behav. Anal.* 36, 21–33. doi: 10.1901/jaba.2003.36-21
- Nesher, P. (1976). Three determinants of difficulty in verbal arithmetic problems. *Educ. Stud. Math.* 7, 369–388. doi: 10.1007/BF00452220
- Nesher, P., and Teubal, E. (1975). Verbal cues as an interfering factor in verbal problem solving. *Educ. Stud. Math.* 6, 41–51. doi: 10.1007/BF00590023
- Nichols, J. (1990). Linguistic diversity and the first settlement of the new world. *Language* 66, 475–521. doi: 10.2307/414609
- Nuerk, H.-C., Geppert, B. E., Van Herten, M., and Willmes, K. (2002). On the impact of different number representations in the number bisection task. *Cortex* 38, 691–715. doi: 10.1016/S0010-9452(08)70038-8
- Nuerk, H.-C., Moeller, K., Klein, E., Willmes, K., and Fischer, M. H. (2011). Extending the mental number line - a review of multi-digit number processing. *J. Psychol.* 219, 3–22.
- Nuerk, H.-C., Moeller, K., and Willmes, K. (2015). "Multi-digit number processing: overview, conceptual clarifications, and language influences," in *The Oxford Handbook of Numerical Cognition*, eds R. C. Kadosh and A. Dowker (Oxford: Oxford University Press). doi: 10.1093/oxfordhnb/9780199642342.013.021
- Nuerk, H.-C., Weger, U., and Willmes, K. (2001). Decade breaks in the mental number line? Putting the tens and units back in different bins. *Cognition* 82, 25–33. doi: 10.1016/S0010-0277(01)00142-1
- Nuerk, H.-C., Weger, U., and Willmes, K. (2005). Language effects in magnitude comparison: small, but not irrelevant. *Brain Lang.* 92, 262–277. doi: 10.1016/j.bandl.2004.06.107
- Nuerk, H.-C., and Willmes, K. (2005). On the magnitude representation of two-digit numbers. *Psychol. Sci.* 42, 52–72.
- OECD. (2010). *PISA 2009 Results. What Students Know and Can Do. Student Performance in Reading, Mathematics and Science*, Vol. 1. Paris: OECD Publishing.
- O'Neill, D. K. (2013). Preschool children's narratives and performance on the peabody individualized achievement test - revised: evidence of a relation between early narrative and later mathematical ability. *First Lang.* 24, 149–183. doi: 10.1177/0142723704043529
- Oostermeijer, M., Boonen, A. J. H., and Jolles, J. (2014). The relation between children's constructive play activities, spatial ability, and mathematical word problem-solving performance: a mediation analysis in sixth-grade students. *Front. Psychol.* 5:782. doi: 10.3389/fpsyg.2014.00782
- Orrantia, J., Rodriguez, L., and Vicente, S. (2010). Automatic activation of addition facts in arithmetic word problems. *Q. J. Exp. Psychol.* 63, 310–319. doi: 10.1080/17470210902903020
- Paige, J. M., and Simon, H. A. (eds). (1966). *Cognitive Processes in Solving Algebra Word Problem*. New York, NY: John Wiley & Sons.
- Pape, S. J. (2003). Compare word problems: consistency hypothesis revisited. *Contemp. Educ. Psychol.* 28, 396–421. doi: 10.1016/S0361-476X(02)00046-2
- Passolunghi, M. C., and Siegel, L. S. (2001). Short-term memory, working memory and inhibitory control in children with difficulties in arithmetic problem solving. *J. Exp. Child Psychol.* 80, 44–57. doi: 10.1006/jecp.2000.2626
- Pixner, S., Kaufmann, L., Moeller, K., Hermanova, V., and Nuerk, H.-C. (2011a). Whorf reloaded: language effects on non-verbal number processing in 1st grade - a trilingual study. *J. Exp. Child Psychol.* 108, 371–382. doi: 10.1016/j.jecp.2010.09.002
- Pixner, S., Zuber, J., Hermanova, V., Kaufmann, L., and Nuerk, H.-C. (2011b). One language, two number-word systems and many problems: numerical cognition in the Czech language. *Res. Dev. Disabil.* 32, 2683–2689. doi: 10.1016/j.ridd.2011.06.004
- Powell, S. R., and Fuchs, L. S. (2014). Does early algebraic reasoning differ as a function of students' difficulty with calculations versus word problems? *Learn. Disabil. Res. Pract.* 29, 106–116. doi: 10.1111/ldrp.12037
- Raduan, I. H. (2010). Error analysis and the corresponding cognitive activities committed by year five primary students in solving mathematical word problems. *Procedia Soc. Behav. Sci.* 2, 3836–3838. doi: 10.1016/j.sbspro.2010.03.600
- Reusser, K. (1988). Problem solving beyond the logic of things: contextual effects on understanding and solving word problems. *Instr. Sci.* 17, 309–338. doi: 10.1007/BF00056219
- Reusser, K. (1993). "Tutoring systems and pedagogical theory: representational tools for understanding, planning and reflection in problem solving," in *Computers as Cognitive Tools*, ed. S. L. S. Derry (Hillsdale, NJ: Lawrence Erlbaum).
- Riley, M. S., Greeno, J. G., and Heller, J. I. (1983). "Development of children's problem-solving ability," in *The Development of Mathematical Thinking*, ed. H. P. Ginsberg (San Diego, CA: Academic Press), 153–196.
- Roettger, T. B., and Domahs, F. (2015). Grammatical number elicits SNARC and MARC effects as a function of task demands. *Q. J. Exp. Psychol.* doi: 10.1080/17470218.2014.979843 [Epub ahead of print].
- Sarnecka, B. W. (2013). On the relation between grammatical number and cardinal numbers in development. *Front. Psychol.* 5:1132. doi: 10.3389/fpsyg.2014.01132
- Schley, D. R., and Fujita, K. (2014). Seeing the math in the story: on how abstraction promotes performance on mathematical word problems. *Soc. Psychol. Personal. Sci.* 5, 953–961. doi: 10.1177/1948550614539519
- Schumacher, R. F., and Fuchs, L. S. (2012). Does understanding relational terminology mediate effects of intervention on compare word problems? *J. Exp. Child Psychol.* 111, 607–628. doi: 10.1016/j.jecp.2011.12.001
- Searle, B. W., Lorton, P. Jr., and Suppes, P. (1974). Structural variables affecting CAI performance on arithmetic word problems of disadvantaged and deaf students. *Educ. Stud. Math.* 5, 371–384. doi: 10.1007/BF01424555

- Shaftel, J., Belton-Kocher, E., Glasnapp, D., and Poggio, J. (2006). The impact of language characteristics in mathematics test items on the performance of english language learners and students with disabilities. *Educ. Assess.* 11, 105–126. doi: 10.1207/s15326977ea1102\_2
- Shaki, S., Fischer, M. H., and Petrusic, W. M. (2009). Reading habits for both words and numbers contribute to the SNARC effect. *Psychon. Bull. Rev.* 16, 328–331. doi: 10.3758/PBR.16.2.328
- Spanos, G., Rhodes, N. C., and Dale, T. C. (1988). “Linguistic features of mathematical problem solving: insights and applications,” in *Linguistic and Cultural Influences on Learning Mathematics*, eds R. R. Cocking and J. P. Mestre (Hillsdale, NJ: Lawrence Erlbaum), 221–240.
- Stacey, K. (2012). “The international assessment of mathematical literacy: PISA 2012 framework and items,” in *Proceedings of the 12th International Congress on Mathematical Education*, Seoul.
- Sternberg, S. (1969). The discovery of processing stages: extensions of Donders’ method. Attention and performance II. *Acta Psychol.* 30, 276–315. doi: 10.1016/0001-6918(69)90055-9
- Swanson, H. L. (2004). Working memory and phonological processing as predictors of children’s mathematical problem solving at different ages. *Mem. Cognit.* 32, 648–661. doi: 10.3758/BF03195856
- Swanson, H. L., Cooney, J. B., and Brock, S. (1993). The influence of working memory and classification ability on children’s word problem solution. *J. Exp. Child Psychol.* 55, 374–395. doi: 10.1006/jecp.1993.1021
- Swanson, H. L., Lussier, C., and Orosco, M. (2013). Effects of cognitive strategy interventions and cognitive moderators on word problem solving in children at risk for problem solving difficulties. *Learn. Disabil. Res. Pract.* 28, 170–183. doi: 10.1111/ldrp.12019
- Terao, A., Koedinger, K. R., Sohn, M. H., Qin, Y., Anderson, J. R., and Carter, C. S. (2004). “An fMRI study of the interplay of symbolic and visuo-spatial systems in mathematical reasoning,” in *Proceedings of the 26th Annual Conference of the Cognitive Science Society*, Chicago, 1327–1332.
- Thevenot, C., Devidal, M., Barrouillet, P., and Fayol, M. (2007). Why does placing the question before an arithmetic word problem improve performance? A situation model account. *Q. J. Exp. Psychol.* 60, 43–56. doi: 10.1080/17470210600587927
- Thevenot, C., and Oakhill, J. (2005). The strategic use of alternative representation in arithmetic word problem solving. *Q. J. Exp. Psychol.* 58, 1311–1323. doi: 10.1080/02724980443000593
- van der Schoot, M., Bakker Arkema, A. H., Horsley, T. M., and Van Lieshout, E. C. D. M. (2009). The consistency effect depends on markedness in less successful but not successful problem solvers: an eye movement study in primary school children. *Contemp. Educ. Psychol.* 34, 58–66. doi: 10.1016/j.cedpsych.2008.07.002
- Van Dijk, T., and Kintsch, W. (1983). *Strategies of Discourse Comprehension*. New York, NY: Academic Press.
- Van Dooren, W., Verschaffel, L., and Onghena, P. (2002). The impact of pre-service teachers’ content knowledge on their evaluations of students’ strategies for solving arithmetic and algebra word problems. *J. Res. Math. Educ.* 33, 319–351. doi: 10.2307/4149957
- Vergnaud, G. (2009). The theory of conceptual fields. *Hum. Dev.* 5, 83–94. doi: 10.1159/000202727
- Verguts, T., and Fias, W. (2005). Interacting neighbors: a connectionist model of retrieval in single-digit multiplication. *Mem. Cogn.* 33, 1–16. doi: 10.3758/BF03195293
- Verschaffel, L. (1994). Using retelling data to study elementary school children’s representations and solutions of compare problems. *J. Res. Math. Educ.* 25, 141–165. doi: 10.2307/749506
- Verschaffel, L., and De Corte, E. (1990). “Do non-semantic factors also influence the solution process of addition and subtraction word problems?” in *Learning and Instruction. European Research in an International Context*, eds H. Mandl, E. De Corte, N. Bennett, and N. H. H. F. Friedrich (Oxford: Pergamon), 415–429.
- Verschaffel, L., De Corte, E., and Borghart, I. (1997). Pre-Service teachers conceptions and beliefs about the role of real-world knowledge in mathematical modelling of school word problems. *Learn. Instr.* 7, 339–359. doi: 10.1016/S0959-4752(97)00008-X
- Verschaffel, L., De Corte, E., and Pauwels, A. (1992). Solving compare problems: an eye movement test of Lewis and Mayer’s consistency hypothesis. *J. Educ. Psychol.* 84, 85–94. doi: 10.1037/0022-0663.84.1.85
- Verschaffel, L., Greer, B., and De Corte, E. (2000). Making sense of word problems. *Educ. Stud. Math.* 42, 211–213.
- Vicente, S., Orrantia, J., and Verschaffel, L. (2007). Influence of situational and conceptual rewording on word problem solving. *Br. J. Educ. Psychol.* 77, 829–848. doi: 10.1348/000709907X178200
- Wilkins, J. L., Baroody, A. J., and Tiilikainen, S. (2001). Kindergartners’ understanding of additive commutativity within the context of word problems. *J. Exp. Child Psychol.* 79, 23–36. doi: 10.1006/jecp.2000.2580
- Wood, G., Nuerk, H.-C., Moeller, K., Geppert, B., Schnitker, R., Weber, J., et al. (2008). All for one, but not one for all. How multiple number representations are recruited in one numerical task. *Brain Res.* 1187, 154–166. doi: 10.1016/j.brainres.2007.09.094
- Yeap, B. H., Ho, S. Y., Kaur, B., and Lee, N. H. (2005). Children making sense during word problem solving. *Papers from the International Conference “Redesigning Pedagogy: Research, policy and practice,”* Singapore. Available at: <http://math.nie.edu.sg/T3/downloads/rprpp.pdf> (accessed March 2015).
- Yeap, B. H., and Kaur, B. (2001). “Semantic characteristics that make arithmetic word problems difficult,” in *Proceedings of the 24th Annual Conference of the Mathematics Education Research Group of Australasia Incorporated (MERGA) on “Numeracy and Beyond,”* Australia.
- Yee, F. P., and Lee, K. P. (1997). School word problems and stereotyped thinking. *Teach. Learn.* 18, 73–82.
- Zarnhofer, S., Braunstein, V., Ebner, F., Koschutnig, K., Neuper, C., Ninaus, M., et al. (2013). Individual differences in solving arithmetic word problems. *Behav. Brain Funct.* 9, 28. doi: 10.1186/1744-9081-9-28
- Zuber, J., Pixner, S., Moeller, K., and 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. *J. Exp. Child Psychol.* 102, 60–77. doi: 10.1016/j.jecp.2008.04.003

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Daroczy, Wolska, Meurers and Nuerk. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Intransparent German number words complicate transcoding – a translingual comparison with Japanese

Korbinian Moeller<sup>1,2,3\*</sup>, Julia Zuber<sup>2</sup>, Naoko Olsen<sup>4</sup>, Hans-Christoph Nuerk<sup>1,2,3</sup> and Klaus Willmes<sup>4</sup>

<sup>1</sup> Knowledge Media Research Center, Tuebingen, Germany, <sup>2</sup> Department of Psychology, University of Tuebingen, Tuebingen, Germany, <sup>3</sup> LEAD Graduate School, University of Tuebingen, Tuebingen, Germany, <sup>4</sup> Section Neuropsychology, Department of Neurology, RWTH Aachen University, Tuebingen, Germany

## OPEN ACCESS

### Edited by:

Yvette Renee Harris,  
Miami University, USA

### Reviewed by:

Barbara W. Sarnecka,  
University of California, Irvine, USA  
Angels Colome,  
University of Barcelona, Spain

### \*Correspondence:

Korbinian Moeller,  
Knowledge Media Research Center,  
Schleichstrasse 6, 72076 Tuebingen,  
Germany  
k.moeller@iwm-kmrc.de

### Specialty section:

This article was submitted to  
Developmental Psychology,  
a section of the journal  
Frontiers in Psychology

**Received:** 27 November 2014

**Accepted:** 18 May 2015

**Published:** 11 June 2015

### Citation:

Moeller K, Zuber J, Olsen N,  
Nuerk H-C and Willmes K (2015)  
Intransparent German number words  
complicate transcoding –  
a translingual comparison with  
Japanese.  
Front. Psychol. 6:740.  
doi: 10.3389/fpsyg.2015.00740

Superior early numerical competencies of children in several Asian countries have (amongst others) been attributed to the higher transparency of their number word systems. Here, we directly investigated this claim by evaluating whether Japanese children's transcoding performance when writing numbers to dictation (e.g., "twenty five" → 25) was less error prone than that of German-speaking children – both in general as well as when considering language-specific attributes of the German number word system such as the inversion property, in particular. In line with this hypothesis we observed that German-speaking children committed more transcoding errors in general than their Japanese peers. Moreover, their error pattern reflected the specific inversion intransparency of the German number-word system. Inversion errors in transcoding represented the most prominent error category in German-speaking children, but were almost absent in Japanese-speaking children. We conclude that the less transparent German number-word system complicates the acquisition of the correspondence between symbolic Arabic numbers and their respective verbal number words.

**Keywords:** transcoding, German, Japanese, number-word system

## Introduction

Recent years have witnessed increasing research interest in the impact of specific language properties on numerical development. A large proportion of these studies focused on the comparison of Western (mostly European and American English) and Asian (mostly Korean, Japanese, and Chinese) children's performance in mathematics. Contrasting these different languages and their cultural backgrounds revealed impressive differences in favor of children from those Asian countries (e.g., Stevenson et al., 1985; Stigler et al., 1987; Miura et al., 1999). For example, Geary et al. (1992) found that Chinese first graders were faster and more accurate in addition tasks than matched US children. Similarly, superiority in subtraction performance of Korean children over US children was reported (Song and Ginsburg, 1987; Fuson and Kwon, 1992). However, these differences are not restricted to more complex mathematical tasks like mental calculation. Even in basic numerical tasks such as counting or place-value understanding differences favoring Asian children were observed (mostly Chinese children: Miura et al., 1988; Miller et al., 1995). Several reasons have been proposed to explain

this superiority of children in several Asian countries. On the one hand, various cultural differences have been mentioned, including variations in home experiences (e.g., greater parental expectations; Song and Ginsburg, 1987; Stevenson and Lee, 1990) as well as differences of educational systems (e.g., quality and quantity of mathematics instruction, rigor, or structure of the mathematics curriculum; Stevenson et al., 1985, 1987; Stigler et al., 1987; Chen and Stevenson, 1989; Hess and Azuma, 1991; Perry et al., 1993; Stevenson and Nerison-Low, 2000). However, it has to be considered that superior performance in basic numerical tasks was already reported before schooling or formal education starts (e.g., Stevenson et al., 1987), questioning the influence of schooling as the only relevant factor (see Miller et al., 2005 for a review).

As a consequence, it was also suggested that specific language characteristics such as the higher transparency of the number word systems of east-Asian languages, such as Japanese and Chinese, and their consistent reflection of the place-value structure of the Arabic number system might also have an impact on mathematics performance (Miura et al., 1988; Geary et al., 1992; Miller et al., 1995; Miura and Okamoto, 2003; see Ng and Rao, 2010 for a recent review; but see Ackerman, 1988 for limitations of this view).

In our view, two approaches may help to differentiate influences of language from those of culture more generally. First, language influences may be examined within the same culture and educational system. For instance Imbo et al. (2014) compared transcoding performance in Dutch- and French-speaking children in Belgium and observed advantages for French-speaking children (see also Dowker and Lloyd, 2005; Dowker et al., 2008; Colomé et al., 2010; Pixner et al., 2011; Salillas and Carreiras, 2014 for studies following this approach). Second, one might aim at considering the specificities of certain languages. Rather than just showing that Japanese or Chinese children are somehow and/or generally better in basic numerical and/or arithmetic tasks than their Western (e.g., German or English) peers, it would be instructive to show that they perform specifically better on those stimuli within the same task, for which the transparency of their number word system gives them a particular advantage. Vice versa, for stimuli for which the Japanese or Chinese number word system provides no particular advantage, differences should be smaller or non-existent at all. Importantly, general cultural differences cannot easily explain such differential effects, when differences between groups can be observed exclusively or predominantly for stimuli which differ with respect to specific attributes of the respective language systems.

In the current study we pursued this second rationale by investigating differences between Japanese and German children attending first grade of primary school regarding basic numerical abilities of transcoding and thus place-value processing. In the following, we will first briefly describe recent evidence concerning language influences on number processing before elaborating on the specific differences between the Japanese and German number word systems from which we derive our hypotheses.

## Language Influence on Numerical Performance

In general, the idea of a language-specific influence on numerical cognition is not new. Quite a few studies found that language-specific features influence performance in numerical tasks. For instance, Colomé et al. (2010) investigated influences of differences in number word formation between Spanish and Basque on adults' addition performance. While Spanish number words reflect the base-10 structure of the Arabic number system, some Basque number words reflect a vigesimal base-20 structure. This means that number words are formed by combining multiples of 20 and units or teens (e.g., "36" is spoken as "hogeita hamasei" literally meaning "twenty and sixteen"). The authors observed that only Basque participants solved additions faster when they were presented as a multiple of 20 and a teen (e.g.,  $20 + 16$ ) as compared to problems with the same results but emphasizing a base-10 structure composition [e.g.,  $26 + 10$ , see also Salillas and Carreiras (2014) for influences of Basque number words on number processing]. Moreover, language-specific influences on numerical performance have also been reported for children. For example, Seron and Fayol (1994) observed language influences comparing French- and Belgian-French-speaking children. In Belgium, decade structures like 70 and 90 are composed regularly ["septante" ("seventy") and "nonante" ("ninety")], whereas in French they are irregular ["soixante-dix" ("sixty-ten") and "quatre-vingt-dix" ("fourty-twenty-ten")]. When children were asked to write down numbers to dictation (e.g., transcoding verbal number words to the corresponding Arabic number), Belgian children committed fewer errors on the respective decades than French children. Moreover, for French-speaking children error types clearly reflected the verbal lexical primitives used to express these decades. For instance, "quatre-vingt-dix-sept" ("four-twenty-ten-seven," which is the corresponding French number word for  $97 = 4 * 20 + 17$ ) was written as 4217, 42017, or 8017 (see also Krinzinger et al., 2011, for a comparison of French, Dutch, and German; Göbel et al., 2014 for language influences on arithmetic).

Moreover, in several number-word systems (e.g., German, Dutch, Arabic, Maltese, Malagasy, etc., Comrie, 2005) tens and units are uttered in reversed order with respect to their order in Arabic notation (e.g., in German "21" is spoken as "einundzwanzig," i.e., "one-and-twenty" translated literally) – referred to as the inversion property of number words. Interestingly, transcoding performance of German-speaking children was found to be severely influenced by the inversion property of German number words. In fact, about 50% of transcoding errors of German-speaking first-graders were related to inversion (Zuber et al., 2009). In contrast, transcoding studies in languages without inversion (except for teen numbers, e.g., "thirteen" in English) did not specifically report inversion errors (e.g., French: Barrouillet et al., 2004; Camos, 2008; Italian: Power and Dal Martello, 1990, 1997). Different studies replicated this observation (e.g., Imbo et al., 2014, for a comparison of Dutch and French in Belgian children, see also Pixner et al., 2011 for a comparison of inverted and non-inverted number words in Czech). These findings provide

first evidence that transcoding performance may somehow be related to language-specific features. However, those studies were restricted to a comparison among different European cultures.

While there are, to the best of our knowledge, no translational studies directly contrasting transcoding in some Western and Asian number-word systems, there are some studies investigating the understanding of the base-10 place-value structure of the Arabic number system. In a first approach, Miura et al. (1988, 1994; Miura and Okamoto, 1989, 2003) assessed whether Asian (including Chinese, Japanese, and Korean) differed from Western (including French, Swedish, and US) children with regard to their representation of the base-10 place-value structure of the Arabic number system. They asked children to construct various numbers by using base-10-blocks. Indeed, children considered how their specific languages reflect or translate the place-value structure of the Arabic number system into their number words. Miura et al. (1988, 1994) suggested that better performance with regard to base-10 understanding of these Asian children is due to a strong influence of language, namely the more transparent correspondence of number words to the place-value structure of Arabic numbers in the respective languages. However, these findings were questioned in subsequent studies. Towse and Saxton (1997) demonstrated that English-speaking children showed similar base-10 place-value understanding as compared to the Asian samples investigated by Miura et al. (1988, 1994; including Chinese, Japanese, and Korean children) when instructed appropriately.

The current study picks up this argument and evaluates the account of Miura et al. (1988) explicitly. If Japanese children have better place-value understanding of the Arabic number system due to higher transparency of their number word system, they should commit less place-value related errors when transcoding number words into Arabic numbers. In particular, errors related to specific intransparencies in comparison to another number word system without these attributes should be examined. Therefore, the current study is designed to compare Japanese- and German-speaking children's performance in a basic numerical transcoding task. Contrasting children's performance in these two disparate number word systems should provide further insight into the extent to which language influences the acquisition of fundamental numerical abilities.

Before introducing our hypotheses in more detail, the structure of the Japanese and German number word system will be sketched briefly, to outline their peculiarities and their possible impact on number processing.

### Differences between the Japanese and the German Number Word System

Number word systems all over the world can differ in several aspects (e.g., base, order, etc.; Comrie, 2005). In several Asian languages, such as Japanese, the number word systems are very transparent. Japanese children only have to memorize the number names from one to nine and the multipliers "juu" ("ten"), "hyaku" ("hundred"), and "sen" ("thousand"), etc.; larger

numbers are then generated according to a set of rules. Decade names are formed by multiplicative composition, e.g., 40 is "yon-juu" ("four-ten"), larger numbers combine multiplicative and additive composition, e.g., 48 is "yon-juu-hachi" ("four-ten-eight"). So there is a consistent relationship between number words and corresponding digits as well as the multiplier in the place-value structure of the Arabic number system for all multi-digit numbers. In Japanese, the order in which units, tens, hundreds, etc. are named in number words thus follows the corresponding order of Arabic digits in a multi-digit number. However, this is different in some Western languages such as German. Here, the order in which tens and units are uttered is inverted in teens and all other two-digit number words: e.g., 21 is pronounced as "one-and-twenty" ("twenty-one").

Furthermore, in Japanese Arabic digits are named identically in number words irrespective of their position within the number (e.g., 2 → "two"; 20 → "two-ten"). In contrast, Arabic digits correspond to different number words at the tens position as compared to the unit position in German number words (e.g., 2 as the number word "two" vs. 2 in "twenty"). Finally, a third difference refers to the name of the multiplier. In Japanese, the multiplier is explicitly part of the spoken number word. For instance, 40 ( $4 * 10$ ) is spoken "yon-juu" ("four-ten"), and 400 ( $4 * 100$ ) is spoken "yon-hyaku" ("four-hundred"). In German, the multiplier is only transparent from three-digit numbers upward [e.g., 400 → "vier hundert" ("four-hundred")], but intransparent for two-digit numbers [e.g., 40 → "vierzig" ("fourty") instead of "vier-zehn" ("four-ten") as in Japanese, see Ng and Rao, 2010 for a review on the influence of Asian number word systems].

However, there are also some intransparencies common to both languages. These concern the role of the digits 0 and 1 in three-digit number words. Both languages do not name "zero" at the tens place (e.g., "207" is "two-hundred and seven" and not "two-hundred-zero-ten-seven"). This intransparency might cause additive composition errors where either zero is left out ("two-hundred and seven" → 27) or the overwriting rule of zeros is ignored ("two-hundred and seven" → 2007). Similarly, "one" is not named at the tens position in both languages ("217" is named as "two-hundred-ten-seven" in Japanese and not "two-hundred-one-ten-seven"). Thus, there is only a multiplier ("ten") for the tens digit, but no value for the digit itself. Therefore, the value of the corresponding Arabic digit cannot be determined from the number-word (e.g., no digit value named in a three-digit number word might as well reflect the value "zero" or "one").

Taken together, these two number word systems differ in several aspects with the Japanese number word system being the more transparent one. If children's errors are related to the specificities of their number word system when they translate one number format to another, this would be an indication that language influences numerical performance. Because the German number word system is rather intransparent compared to the Japanese one due to its inversion property, it is expected that German speaking children commit more errors reflecting their problems with understanding the place-value structure of Arabic

numbers. Generally, this refers to errors violating the syntactic structure of the respective multi-digit number (see Materials and Methods section for a taxonomy of transcoding errors) such as additive and multiplicative composition errors (e.g., “two-hundred seven” → 2007) as well as inversion errors. As described above there are commonalities and differences between German and Japanese with respect to transparency in additive and multiplicative composition. Nevertheless, because digits correspond to specific number names at the tens position (e.g., 2 → “twenty”) and the fact that the multiplier is not indicated in German number words denoting the decades, we expected more additive and multiplicative transcoding errors for German- as compared to Japanese-speaking children. Importantly, however, the inversion property and associated inversion errors are of highest interest in this study because there is no number word inversion in Japanese at all. In Contrast, German children’s transcoding errors have been found to be inversion related in 50% of the cases [i.e., “twenty-five” (spoken as “five and twenty”) → 52, Zuber et al., 2009], thereby reflecting a number-word specific intransparency. As no inversion of tens and units is present in the Japanese number word system, no such errors should occur in Japanese-speaking children. Thus, given an influence of language on performance, error rates should not only differ in general, but should also be differentially related to specific attributes of the number word structure of the respective languages.

## Materials and Methods

### Participants

In total, 40 children participated in the study. Twenty German-speaking children (10 girls), were recruited from a German elementary school, mean age was 7.32 years (SD = 0.36; range 6 years 7 months to 7 years 8 months). All children spoke German as their native language, none of them had been noted for having specific difficulties in mathematics or other school problems. Additionally, twenty Japanese children (seven girls) were recruited from a Japanese elementary school in Germany. Their mean age was 7.27 years (SD = 0.36; range 6 years 5 months to 7 years 7 months). Japanese schools in Germany follow the Japanese curricula and teaching is exclusively in Japanese. Moreover, all children’s parents were both native speakers of Japanese, and only Japanese was spoken at home. Additionally, Japanese children did not speak any German nor had they encountered German numbers, yet. According to the respective school curricula the number of mathematics classes is equal for both language groups. By the end of first grade, all children should know the numbers up to 20 and be able to perform simple additions and subtractions within this range. To furthermore ensure an equal level of education, both groups were tested toward the end of the academic year, this means German-speaking children at the end of May and Japanese-speaking children in February, because the Japanese academic year ends in March.

The study was approved by the local school authorities and carried out in line with the latest version of the Declaration of

Helsinki. Written informed consent was obtained from parents of all participating children prior to the study.

### Tasks and Stimuli

The *transcoding* task consisted of 67 stimuli (i.e., 9 single-digit, 36 two-digit, and 22 three-digit numbers), incorporating all lexical primitives and different syntactic structures. Children had to write them down as Arabic numbers to dictation. Numerical structures not yet learned at school (i.e., three-digit numbers) were presented in order to assess whether children were able to apply and generalize rules they had already learned on simpler forms (see Byrge et al., 2014 for kindergartner’s writing down three-digit numbers). The order of the stimuli was randomly assigned, but the task always started with a one-digit number.

There was also a block of items, in which children had to read aloud Arabic numbers. However, as the results did not differ substantially between these two conditions and the error analysis of the reading aloud condition is less discriminating (e.g., no child would name 324 as “thirty thousand twenty four” but when instructed to write down “three hundred twenty four” in Arabic notation 30024 is a quite common error) this article focuses on the results of the writing to dictation condition.

### Procedure

Children were tested individually in a quiet room during school hours in one-on-one sessions. Children had to write down numbers to dictation on a blank sheet, one below another. No feedback was given as to the correctness of the results. The critical 67 trials were preceded by two practice trials to familiarize children with the task.

### Transcoding Error Analysis

Errors were categorized according to the taxonomy used in Zuber et al. (2009; extended and slightly modified from Deloche and Seron, 1982). This categorization is used because it allows classification of inversion errors; moreover, it is kept as general as possible to enable its use in a variety of languages.

In general, this categorization distinguishes *lexical* from *syntactic* errors (following Deloche and Seron, 1982, 1987). Lexical errors concerned the substitution of one (or more) lexical elements by another one with no modification of the syntactic structure. This error category was subdivided into lexical value errors being (a) zero dependent, e.g., “eighty” → 81; or (b) zero independent, e.g., “thirty-four” → 35, and (c) lexical class errors, where the primitive itself is correct but its class is not (e.g., “eighty” → 18). Lexical errors that could not be classified into one of these categories were coded as (d) other lexical errors.

Errors were classified to be *syntactic* when they altered the syntactic structure of the produced numeral compared to the target form. This could either be due to violations of the (a) additive composition rule e.g., “three hundred twenty” → 30020 (when twenty is appended in the composition rather than added) and (b) multiplicative composition rule, e.g., “three hundred” → 3100 (when 100 is appended in the composition rather than multiplied. Further, (c) inversion errors were also categorized as syntactic errors as they mirror the understanding of syntactic

rules (i.e., place-value structuring). Inversion errors could either be due to disregard of inversion meaning that the to-be-inverted digits were produced in the wrong order [e.g., “twenty-five” (“five and twenty”) → 52], or reflecting wrong application of inversion, this means, when hearing “three hundred,” children may wrongly apply the inversion rule (e.g., “three hundred” → 103) reflecting an overgeneralization of this rule. Again, errors that could not be classified into these subcategories were coded as (d) other syntactic errors.

Eventually, errors including both wrong lexical elements and incorrect syntactic structures were coded as *combination errors*.

Finally, transcoding errors that could not be classified as belonging to one of the categories specified above were coded as *other errors*.

## Results

Inferential statistics were conducted on arcsine-transformed error proportions to approximate normal distributions. In case the sphericity assumption was violated, the original degrees of freedom together with the respective Greenhouse–Geisser coefficients (GGs) are reported. One German-speaking child was excluded from further analyses because 24 of the 25 transcoding errors of this child were non-responses. For the remaining participants there were 2.2% non-responses in German speaking children and 0.3% in Japanese children, which were not included in the analyses.

### Overall Error Categories

To examine whether absolute error rates differed between the languages a 2 (language) × 4 (error categories: lexical, syntactic, combination errors, others) ANOVA was conducted. The ANOVA revealed significant main effects for both factors [language:  $F(1,37) = 31.72, p < 0.001, \eta_p^2 = 0.46$ ; error category:  $F(3,111) = 38.66, p < 0.001, \eta_p^2 = 0.51, GG = 0.86$ ] and a significant interaction [ $F(3,111) = 13.31, p < 0.001, \eta_p^2 = 0.27, GG = 0.86$ ]. German-speaking children committed reliably more

errors in general than their Japanese-speaking counterparts (7.2 vs. 1.4%, respectively). Additionally, the frequency of error categories differed significantly: pairwise contrasts indicated that syntactic errors were reliably more frequent than all other error categories (10.0%; all  $p < 0.001$ , Bonferroni-corrected) whereas there were no reliable differences between the remaining error categories (lexical errors: 1.7%, combination errors: 2.9%, other errors: 2.1%; all  $p > 0.9$ ). The reliable two-way interaction indicated that languages differed reliably for the respective profile of error categories (see **Table 1A**). To evaluate our hypothesis that differences should be most pronounced for syntactic errors, we conducted three additional two-way ANOVAs with the factors language group and error category in which the latter reflected all possible pairwise combinations of the syntactic and one of the other error categories (i.e., syntactic vs. lexical errors; syntactic vs. combination errors, syntactic vs. other errors). To account for influences of multiple testing we reduced the alpha level accordingly (significant when  $p < 0.05/3 = 0.017$ ). The ANOVAs consistently revealed reliable interactions of language group and error categories for syntactic vs. lexical errors [ $F(1,37) = 29.65, p < 0.001, \eta_p^2 = 0.45$ ], syntactic vs. other errors [ $F(1,37) = 22.87, p < 0.001, \eta_p^2 = 0.38$ ] as well as the interaction of language group and syntactic vs. combination errors [ $F(1,37) = 5.78, p = 0.021, \eta_p^2 = 0.14$ ]. Importantly, these interactions indicated language differences for syntactic errors (14.0%; German: 17.4% vs. Japanese: 3.4%) to be more pronounced than those for lexical (1.8%; German: 2.7% vs. Japanese: 0.9%), other errors (2.0%; German: 2.3% vs. Japanese: 0.3%), and combination errors (6.0%; German: 6.1% vs. Japanese: 0.1%). Furthermore, simple effects indicated that language differences were reliable for all error categories with German-speaking children consistently committing more transcoding errors [syntactic:  $t(37) = 6.41, p < 0.001$ ; combined [ $t(37) = 4.51, p < 0.001$ ; lexical errors:  $t(37) = 2.36, p < 0.05$ ; other errors:  $t(37) = 2.44, p < 0.05$ ]. In sum, this corroborated our hypothesis that language differences should be most pronounced for syntactic errors, as these include inversion errors that should be specific to German-speaking children.

**TABLE 1 | Overview of absolute error rates for all error categories (A) as well as absolute and relative error rates for subcategories of syntactic (B) and lexical errors (C) separated for German- and Japanese-speaking children, SEM given in parentheses.**

Error categories	German		Japanese	
<b>(A) Absolute overall error rates</b>				
Syntactic errors	17.4% (1.9)		3.4% (1.9)	
Lexical errors	2.7% (0.7)		0.9% (0.7)	
Combination errors	6.1% (1.5)		0.1% (1.4)	
Other errors	2.3% (0.7)		0.3% (0.7)	
<b>(B) Subcategories syntactic errors</b>				
	Absolute	Relative	Absolute	Relative
Additive composition errors	7.2% (1.7)	33.3% (8.6)	2.9% (6.0)	75.8% (12.4)
Multiplicative composition errors	1.3% (2.0)	6.2% (2.0)	0.2% (1.0)	1.8% (2.8)
Inversion errors	8.6% (6.7)	58.5% (8.8)	0.3% (0.7)	22.2% (12.7)
<b>(C) Subcategories lexical errors</b>				
Lexical value errors	0.5% (0.3)	9.7% (8.4)	0.1% (0.1)	20.0% (13.0)
Lexical value errors incl. zero	0.6% (0.5)	13.3% (10.9)	0.4% (0.3)	36.0% (16.9)
Lexical class errors	0.9% (0.4)	31.7% (10.5)	0.1% (0.1)	4.0% (16.2)

## Error Subcategories

To investigate whether these different error distributions were indeed specifically related to number word attributes the *absolute* and *relative* error frequencies of the subcategories of syntactic and lexical errors per child were evaluated in more detail.

We did not consider combination errors and other errors here because we had no specific hypothesis for language effects on the latter. Regarding combination errors, Japanese-speaking children did not commit any error in three of the four categories of combination errors we observed (i.e., combination of lexical and inversion errors, lexical, syntactic and inversion errors, as well as syntactic and inversion errors). For the remaining combination of lexical and syntactic errors there was only one Japanese child who committed one such error. Therefore, we refrained from further analyzing frequencies of error subcategories of combination errors.

To allow applicability of ANOVA methods for the patterns of *relative* error frequencies, we excluded one subcategory of errors each from the analyses to avoid complete dependency among error categories in that they would always sum up to 100%. We eliminated the categories 'other syntactic errors' and 'other lexical errors' because they were of only marginal theoretical interest. Thus, relative error frequencies do not add up to 100%. To keep the analyses of absolute and relative error rates comparable we also excluded the categories 'other syntactic errors' and 'other lexical errors' when analyzing the absolute rates of error subcategories. This means that the analysis for syntactic errors discerned the subcategories inversion errors, as well as additive and multiplicative composition errors. On the other hand, the analysis for lexical errors discerned the subcategories of lexical class errors, lexical value errors not including and lexical value errors including zero.

Moreover, for the analysis of relative frequencies all children who did not commit at least one error were excluded from analyses since they do not contribute to potential differentiation between error subcategories. For syntactic errors, this affected 11 of the Japanese-speaking children and no German-speaking child. For lexical errors this affected 15 Japanese- and 7 German-speaking children. Because of this considerable reduction of sample sizes and the generally low frequencies of lexical errors the results for the specific evaluation of lexical error subcategories need to be treated with caution.

## Syntactic Errors

### Absolute error rates

The 2 (language group)  $\times$  3 (error subcategory: additive composition, multiplicative composition, inversion) ANOVA revealed a reliable main effect of language group [ $F(1,37) = 37.34$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.34$ ] indicating that German-speaking children committed significantly more syntactic transcoding errors across all subcategories than their Japanese-speaking peers (5.7% vs. 1.1%, respectively). Additionally, the main effect of error subcategory was reliable [ $F(2,74) = 10.47$ ,  $p < 0.01$ ,  $GG = 0.69$ ;  $\eta_p^2 = 0.22$ ] suggesting that error rates were not distributed equally across syntactic error subcategories. Pairwise comparisons showed that multiplicative composition errors were reliably less frequent (0.8%, both  $p < 0.05$ ,

Bonferroni corrected) than both additive composition (5.1%) and inversion errors (4.4%). This error pattern was further qualified by language as indicated by the significant interaction between language group and error subcategory [ $F(2,74) = 7.50$ ,  $p < 0.01$ ,  $GG = 0.69$ ,  $\eta_p^2 = 0.17$ , see **Table 1B**]. As we hypothesized that language differences within the category of syntactic errors should be driven by the specifically increased frequencies of inversion errors in German-speaking children, we conducted two additional two-way ANOVAs with the factors language group and error subcategory. The factor error subcategory reflected the pairwise combinations of inversion errors with other syntactic error subcategories (i.e., inversion vs. multiplicative composition errors; inversion vs. additive composition errors). We reduced the alpha level accordingly to control for influences of multiple comparisons (i.e., significant when  $p < 0.05/2 = 0.025$ ). The ANOVAs revealed a marginally reliable interaction of language group and error subcategories inversion vs. additive composition errors [ $F(1,37) = 4.80$ ,  $p < 0.05$ ,  $\eta_p^2 = 0.12$ ] whereas the interaction for error subcategories inversion vs. multiplicative composition errors was highly significant [ $F(1,37) = 30.44$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.45$ ]. Importantly, the former interaction indicated that the language difference for inversion errors (8.4%; German: 8.6% vs. Japanese: 0.2%) tended to be more pronounced than that for additive composition errors (4.3%; German: 7.2% vs. Japanese: 2.9%). Moreover, the language difference was significantly more pronounced for inversion errors than for multiplicative composition errors (1.1%, German: 1.3% vs. Japanese: 0.2%).

Tests for simple effects indicated that absolute frequencies of error subcategories were significantly higher for German-speaking children for all error subcategories [inversion errors: 8.6% vs. 0.2%,  $t(37) = 8.55$ ,  $p < 0.001$ ; additive composition errors: 7.2% vs. 3.0%,  $t(37) = 2.04$ ,  $p < 0.05$ ; multiplicative composition errors: 1.3% vs. 0.2%,  $t(37) = 2.62$ ,  $p < 0.05$ ].

### Relative error rates

As to be expected, the 2 (language group)  $\times$  3 (error subcategory: additive composition, multiplicative composition, inversion) ANOVA on relative error frequencies revealed no significant main effect of language group [ $F(1,27) < 1$ ]. However, the main effect of error subcategory was reliable [ $F(2,52) = 10.96$ ,  $p < 0.001$ ;  $\eta_p^2 = 0.30$ ,  $GG = 0.57$ ] suggesting that syntactic error subcategories were not distributed equally. Pairwise comparisons showed that multiplicative composition errors were reliably less frequent (4.0%, both  $p < 0.05$ , Bonferroni corrected) than both additive composition (54.6%) and inversion errors (40.4%). Importantly, however, this error pattern was qualified by language as indicated by the reliable interaction between language group and error subcategory [ $F(2,52) = 7.59$ ,  $p < 0.01$ ,  $\eta_p^2 = 0.23$ ,  $GG = 0.57$ , see **Table 1B**]. To evaluate whether language differences within the category of syntactic errors were indeed driven by the specifically increased frequencies of inversion errors in German-speaking children, two additional two-way ANOVAs with the factors language group and error subcategory were carried out. The latter factor reflected the pairwise combinations of inversion errors with other syntactic

error subcategories (i.e., inversion vs. multiplicative composition errors; inversion vs. additive composition errors). To account for influences of multiple testing we reduced the alpha level accordingly (significant when  $p < 0.05/2 = 0.025$ ). The ANOVAs revealed a reliable interaction of language group and error subcategories inversion vs. additive composition errors [ $F(1,26) = 8.00, p < 0.01, \eta_p^2 = 0.24$ ] whereas the interaction with error subcategories inversion vs. multiplicative composition errors was not reliable [ $F(1,26) = 3.12, p = 0.09, \eta_p^2 = 0.11$ ]. Importantly, the former interaction indicated that the language differences for inversion errors (36.3%; German: 58.5% vs. Japanese: 22.2%) were indeed more pronounced than for additive composition errors (−42.6%; German: 33.3% vs. Japanese: 75.8%).

Tests for simple effects substantiated that relative frequencies of the error subcategories were significantly higher for German-speaking children for inversion errors [ $t(26) = 2.56, p < 0.05$ ], whereas Japanese-speaking children committed relatively more additive composition errors [ $t(26) = 3.02, p < 0.01$ ]. There was no reliable difference for multiplicative composition errors [ $t(26) = 1.46, p = 0.16$ ].

Taken together, the frequencies of absolute and relative syntactic error subcategories mirrored the hypothesized language specificities. The inversion property of German number words led to a specific absolute but also relative increase of inversion errors not present for Japanese-speaking children. However, these data also indicate that additive composition errors were relatively more prominent in Japanese-speaking children – even though they were more prominent in absolute terms for German-speaking children.

## Lexical Errors

### Absolute error rates

The 2 (language)  $\times$  3 (lexical errors: zero dependent, zero independent, lexical class) ANOVA revealed no reliable effects of the factors error subcategories [ $F(2,74) < 1$ ] and language group [ $F(1,37) = 2.86, p = 0.10, \eta_p^2 = 0.07, GG = 0.89$ ] nor a significant interaction of these two factors [ $F(2,74) = 1.06, p = 0.35, \eta_p^2 = 0.03, GG = 0.89$ , see **Table 1C**].

Simple effects revealed that German-speaking children committed significantly more lexical class errors than Japanese-speaking children [0.9 vs. 0.07%,  $t(37) = 2.04, p < 0.05$ ]. In contrast, there were no reliable language differences for zero-independent [0.5 vs. 0.07%, respectively,  $t(37) = 1.30, p = 0.21$ ] and zero-dependent errors [0.6% vs. 0.4%,  $t(37) = 0.19, p = 0.85$ ].

### Relative error rates

The 2 (language)  $\times$  3 (lexical errors: zero dependent, zero independent, lexical class) ANOVA neither revealed reliable effects of the factors error subcategories [ $F(2,30) < 1$ ] and language group [ $F(1,15) < 1$ ] nor an interaction of these two factors [ $F(2,30) = 1.48, p = 0.24, \eta_p^2 = 0.09$ , see **Table 1C**].

In summary, this pattern is in line with our specificity hypothesis that language differences should be most prominent for syntactic error categories reflecting differences of the number word systems compared.

## Discussion

The aim of this study was to investigate influences of language on numerical development by means of contrasting German- and Japanese-speaking children's transcoding performance by the end of first grade. We were particularly interested in whether there were only general differences in the overall performance level or rather specific error patterns for the two language groups, reflecting the specific intransparencies of the respective number word systems. In particular, we expected inversion errors to be more prominent in German-speaking children.

In line with our expectations we observed strong indications of language influences on transcoding performance. First, German-speaking children, who have to learn the less transparent number word system, committed reliably more transcoding errors in general. However, and more importantly, more fine-grained analyses corroborated our more specific hypothesis that the distribution across error-types should not be arbitrary, but reflect the specificities of the respective number word systems. German-speaking children showed higher absolute rates of syntactic transcoding errors in general and each subcategory of syntactic errors (i.e., inversion, additive, and multiplicative composition) in particular. This reflects less precise overall understanding of the composition of multi-digit numbers out of their single-digit components in German-speaking children. Additionally, within the category of syntactic transcoding errors consideration of absolute and relative error rates indicated that inversion errors were not only the most prominent syntactic error subcategory in German-speaking children but also reliably more prominent than in Japanese-speaking children. The difficulty arising from the inversion property of German number words is further illustrated by the fact that inversion errors were not restricted to errors associated with the order of tens and units and thus the to-be-inverted digits (e.g., “twenty five”  $\rightarrow$  52). Instead, about 25% of inversion errors in German-speaking children reflected an overgeneralization of the inversion rule to hundreds (e.g., “nine hundred”  $\rightarrow$  109). No such error was committed by Japanese-speaking children. This clearly indicated the influence of the inversion property (i.e., the inverted order in which tens and units are named in number words) as a particular language attribute, which is present in German but not in Japanese, on children's place-value understanding. In sum, these data clearly corroborate the hypothesis that language influences numerical abilities. This point and possible reasons for the higher specific and unspecific error rates in German will be discussed in the following.

Essentially, about half of the errors of German-speaking children were related to the inconsistency of inversion, whereas hardly any inversion errors were committed by Japanese-speaking children. Therefore, the interpretation of these results is straightforward. As there is no inversion in Japanese, almost none of these errors occurred; whereas, once inversion is present, the error distribution reflects this intransparency of the number word system. This is in line with the results of Pixner et al. (2011) who investigated transcoding in Czech-speaking children. In Czech both non-inverted and inverted number words for two-digit numbers are used commonly.

Thus, Pixner et al. (2011) were able to directly evaluate the influence of inversion on transcoding performance within the same children. Similar to the present results, Pixner et al. (2011) observed that Czech children committed inversion related transcoding errors only when dictated number words were in the inverted format. However, transcoding errors of German-speaking children were not related exclusively to the specific attribute of inversion in German number words. Instead, error frequencies seemed to reflect the generally higher intransparency of the German number word system with regard to the reflection of the place-value structure of the Arabic number system: absolute frequencies of all subcategories of syntactic (place-value) errors were higher for German-speaking children.

In this respect, it is important to note that the few errors observed in Japanese children were often related to the only intransparency of the Japanese number-word system, such as the missing digit value for “one” (“one” is not named in the decade position, e.g., 217 → “two-hundred-ten-seven” and not “two-hundred-one-ten-seven”) or the missing digit and multiplier values for zero (e.g., “207” is “two-hundred and seven” and not “two-hundred-zero-ten-seven”). These intransparencies are related to additive composition. Accordingly, additive composition transcoding errors had a higher relative frequency in Japanese-speaking than in German-speaking children (even though German-speaking children committed more additive composition errors in absolute terms). Indeed, when examining the errors of Japanese children, it appears that almost all errors were related to the inconsistency of additive composition, whereas there were much fewer errors in all other error categories. Similarly, German speaking children’s additive composition errors constitute the major error subgroup besides inversion errors. Taken together, this supports the hypothesis of a language-specific influence of number word formation on transcoding performance.

However, when comparing error rates for additive composition rules between the two languages, one might wonder, why the absolute error rates of German-speaking children were about four times higher than those of Japanese-speaking children, even though they reflect the same additive composition principle. To account for these findings, the cognitive processes necessary for transcoding should be considered. Generally, the present pattern of results indicated that more transparent number word structures are less error prone, when children have to transcode numbers. But, additionally, another process might be involved as well. Because the structure of Japanese number words is simpler, it may require less working memory (WM) capacity to correctly transcode numbers. Indeed, WM was observed to be an influencing factor in several studies. For instance, Barrouillet et al. (2004) found WM to reliably predict transcoding performance (see also Camos, 2008). Moreover, Zuber et al. (2009) found that WM capacity was specifically important for transcoding in a language with inversion (see also Imbo et al., 2014). Therefore, one might speculate that an intransparent number word system requires more WM capacity and might therefore be more error prone in general. In contrast, a more transparent number word structure like the

Japanese would require less WM capacity and may thus be less susceptible to WM capacity limitations. In this respect, WM capacity limitations in children may be partially responsible for our finding that German children committed more additive composition errors than Japanese children, even though the same principle has to be applied in the two languages.

Although this study revealed reliable influences of language on children’s numerical performance, one cannot exclude the possibility of other factors entirely. It should be acknowledged that even if teaching curricula were the same in both groups, school and home-related factors might have influenced children’s performance as well. In Japan mathematics performance is considered more important than in Western cultures (e.g., Stevenson and Lee, 1990) and children are trained and supported to a greater extent by parents and teachers (e.g., Song and Ginsburg, 1987; Stevenson and Lee, 1990). However, for first grade children, our data indicated that these factors do not seem to be the only ones to influence numerical performance because they can only account for better overall performance of Japanese children. Yet, we also observed specific differences in the distribution of absolute and relative error patterns of the two language groups, which corresponded very closely to the specificities of the two number word systems: German children did not only produced consistently more errors in absolute terms, but also showed higher absolute and relative rates of errors specifically related to the particular intransparencies of their number-word system regarding place-value coding (i.e., the inversion property). These specific effects cannot be explained by an account stressing general differences in learning, education or culture.

Finally, the impact of these language-specific influences on the acquisition of more complex numerical and arithmetical skills has to be considered. There is accumulating evidence that the understanding of basic numerical concepts including the place-value structure of the Arabic number system influences basic numerical (e.g., Holloway and Ansari, 2009; Moeller et al., 2009, 2015; Helmreich et al., 2011; Pixner et al., 2011) but also arithmetic performance (e.g., Levine et al., 1992; Kaufmann et al., 2003; Booth and Siegler, 2008; Göbel et al., 2014). On a very basic level, Cankaya et al. (2014) observed that the regular and transparent Turkish number word structure led to faster acquisition of counting principles and thus better counting performance in Turkish-speaking kindergartners (but see Vasilyeva et al., 2015 for a diverging account). For primary school children Moeller et al. (2011) found specific longitudinal influences of early place-value understanding (as assessed by transcoding performance amongst others) on children’s numerical development. The authors observed that children who committed more inversion-related transcoding errors at the end of grade 1 not only showed poorer addition performance at the end of grade 3 but also had particular difficulties solving addition problems requiring a carry and thus posing increased demands on their place-value understanding. This is further corroborated by data of Moura et al. (2013), who found that children with mathematical difficulties in middle grades of primary school had particular problems acquiring the syntactic transcoding rules allowing for

correct place-value coding of multi-digit numbers. Additionally, Imbo et al. (2014) observed that Dutch second graders who experienced transcoding problems were also found to achieve generally worse in math as indicated by their grades.

Given this strong influence of language on the understanding of the place-value structure of the Arabic number system and the observed relation of the latter with mathematics performance more generally, the present results might also be informative with regard to the repeatedly observed performance difference in mathematical achievement between Western and Asian children (e.g., Fuson and Kwon, 1992). The present data suggest that it is more demanding for children to successfully acquire the relation between symbolic Arabic numbers and number words and thus to acquire place-value understanding in languages with intransparent number word systems. However, the studies described above indicated that such basic place-value understanding is predictive for successful acquisition of more complex arithmetic and mathematical competencies (e.g., Moeller et al., 2011 for the specific case of transcoding). Considering this state of affairs, the implications for teaching are straightforward: for children having to learn an intransparent and complex number word system it might be particularly important to teach and train the correspondence of the Arabic number system and number words more intensively, until children successfully master this link (e.g., Link et al., 2014).

## References

- Ackerman, P. L. (1988). Determinants of individual differences during skill acquisition: cognitive abilities and information processing. *J. Exp. Psychol. Gen.* 117, 288–318. doi: 10.1037/0096-3445.117.3.288
- Barrouillet, P., Camos, V., Perruchet, P., and Seron, X. (2004). ADAPT: a developmental, asemantic and procedural model for transcoding from verbal to Arabic numerals. *Psychol. Rev.* 111, 368–394. doi: 10.1037/0033-295X.111.2.368
- Booth, J., and Siegler, R. (2008). Numerical magnitude representations influence arithmetic learning. *Child Dev.* 79, 1016–1031. doi: 10.1111/j.1467-8624.2008.01173.x
- Byrge, L., Smith, L. B., and Mix, K. S. (2014). Beginnings of place value: how preschoolers write three-digit numbers. *Child Dev.* 85, 437–443. doi: 10.1111/cdev.12162
- Camos, V. (2008). Low working memory capacity impedes both efficiency and learning of number transcoding in children. *J. Exp. Child Psychol.* 99, 37–57. doi: 10.1016/j.jecp.2007.06.006
- Cankaya, O., LeFevre, J. A., and Dunbar, K. (2014). The role of number naming systems and numeracy experiences in children's rote counting: evidence from Turkish and Canadian children. *Learn. Individ. Differ.* 32, 238–245. doi: 10.1016/j.lindif.2014.03.016
- Chen, C. S., and Stevenson, H. W. (1989). Homework: a cross-cultural examination. *Child Dev.* 60, 551–561. doi: 10.2307/1130721
- Colomé, A., Laka, I., and Sebastiaen-Galles, N. (2010). Language effects in addition: how you say it counts. *Q. J. Exp. Psychol.* 63, 965–983. doi: 10.1080/17470210903134377
- Comrie, B. (2005). "Endangered numeral systems," in *Bedrohte Vielfalt: Aspekte des Sprach(en)tods* [Endangered Diversity: Aspects of Language Death], eds J. Wohlgenuth and T. Dirksmeyer (Berlin: Weißensee Verlag).
- Deloche, G., and Seron, X. (1982). From one to 1: an analysis of a transcoding process by means of neuropsychological data. *Cognition* 12, 119–149. doi: 10.1016/0010-0277(82)90009-9
- Deloche, G., and Seron, X. (1987). "Numerical transcoding: a general production model," in *Mathematical Disabilities: A Cognitive Neuropsychological*

## Conclusion

This study showed that German-speaking children were outperformed by Japanese-speaking children not only with respect to overall transcoding performance, but also experienced a particular disadvantage related to specific intransparencies of the German number word system. German-speaking children showed higher absolute error rates in general but also higher absolute and relative error rates specifically reflecting the inversion property of the German number word system. Such a differential performance pattern cannot be explained easily by general cultural accounts emphasizing the role of different learning cultures and/or education. Instead, these results are well in line with language accounts (Miller et al., 2005 for a review) suggesting that transcoding performance should be affected most where the respective number word system is most intransparent.

From this we conclude that the intransparency of the German number word system hampers fast and accurate acquisition of the correspondence between symbolic Arabic numbers and their verbal number names, while the transparency of the Japanese number word system leaves Japanese-speaking children at a considerable advantage. In sum, a better understanding of the difficulties imposed by the specificities of a particular number word system may help to corroborate transcoding skills and thus children's place-value understanding, which – in turn – has been shown to predict future numerical and arithmetic achievement.

- Perspectivae*, eds G. Deloche and X. Seron (Hillsdale, NJ: Erlbaum), 37–179.
- Dowker, A., Bala, S., and Lloyd, D. (2008). Linguistic influences on mathematical development: how important is the transparency of the counting system. *Philos. Psychol.* 21, 523–538. doi: 10.1080/09515080802285511
- Dowker, A., and Lloyd, D. (2005). "Linguistic influences on numeracy," in *Mathematics in the Primary School*, eds A. Dowker and D. Lloyd (Bangor: School of Education), 20–35.
- Fuson, K., and Kwon, Y. (1992). Korean children's understanding of multidigit addition and subtraction. *Child Dev.* 63, 491–506. doi: 10.2307/1131494
- Geary, D. C., Fan, L., and Bow-Thomas, C. C. (1992). Loci of ability differences comparing children from China and the United States. *Psychol. Sci.* 3, 180–185. doi: 10.1111/j.1467-9280.1992.tb00023.x
- Göbel, S., Moeller, K., Pixner, S., Kaufmann, L., and Nuerk, H.-C. (2014). Language affects symbolic arithmetic in children: the case of number word inversion. *J. Exp. Child Psychol.* 119, 17–25. doi: 10.1016/j.jecp.2013.10.001
- Helmreich, I., Zuber, J., Pixner, S., Kaufmann, L., Nuerk, H.-C., and Moeller, K. (2011). Language effects on children's nonverbal number line estimations. *J. Cross Cult. Psychol.* 42, 598–613. doi: 10.1177/0022022111406026
- Hess, R. D., and Azuma, H. (1991). Cultural support for schooling: contrasts between Japan and the United States. *Educ. Res.* 20, 2–8. doi: 10.3102/0013189X020009002
- Holloway, I. D., and Ansari, D. (2009). Mapping numerical magnitudes onto symbols: the numerical distance effect and individual differences in children's math achievement. *J. Exp. Child Psychol.* 103, 17–29. doi: 10.1016/j.jecp.2008.04.001
- Imbo, I., Vanden Bulcke, C., De Brauwier, J., and Fias, W. (2014). Sixty-four or four-and-sixty? The influence of language and working memory on children's number transcoding. *Front. Psychol.* 5:313. doi: 10.3389/fpsyg.2014.00313
- Kaufmann, L., Handl, P., and Thoeny, B. (2003). Evaluation of a numeracy intervention program focusing on basic numerical knowledge and conceptual knowledge: a pilot study. *J. Learn. Disabil.* 36, 564–573. doi: 10.1177/00222194030360060701

- Krinzinger, H., Gregoire, J., Desoete, A., Kaufmann, L., Nuerk, H. C., and Willmes, K. (2011). Differential language effects on numerical skills in second grade. *J. Cross Cult. Psychol.* 42, 614–629. doi: 10.1177/0022022111406252
- Levine, S., Jordan, N. C., and Huttenlocher, J. (1992). Development of calculation abilities in young children. *J. Exp. Child Psychol.* 53, 72–103. doi: 10.1016/S0022-0965(05)80005-0
- Link, T., Schwarz, E. J., Huber, S., Fischer, U., Nuerk, H. C., Cress, U., et al. (2014). Mathe mit der Matte - Verkörperlichtes Training basisnumerischer Kompetenzen. *Z. Erziehungswiss.* 17, 257–277. doi: 10.1007/s11618-014-0533-2
- Miller, K. F., Kelly, M., and Zhou, X. (2005). “Learning mathematics in China and the United States,” in *Handbook of Mathematical Cognition*, ed. J. I. D. Campbell (New York: Psychology press), 163–178.
- Miller, K. F., Smith, C. M., Zhu, J., and Zhang, H. (1995). Preschool origins of cross-national differences in mathematical competence: the role of number-naming systems. *Psychol. Sci.* 6, 56–60. doi: 10.1111/j.1467-9280.1995.tb00305.x
- Miura, I. T., Kim, C. C., Chang, C.-M., and Okamoto, Y. (1988). Effects of language characteristics on children’s cognitive representation of number: cross-national comparisons. *Child Dev.* 59, 1445–1450. doi: 10.2307/1130659
- Miura, I. T., and Okamoto, Y. (1989). Comparisons of U.S. and Japanese first graders’ cognitive representation of number and understanding of place value. *J. Educ. Psychol.* 81, 109–114. doi: 10.1037/0022-0663.81.1.109
- Miura, I. T., and Okamoto, Y. (2003). “Language supports for mathematics understanding and performance,” in *The Development of Arithmetic Concepts and Skills*, eds A. Baroody and A. Dowker (Mahwah, NJ: Lawrence Erlbaum Associates), 229–242.
- Miura, I. T., Okamoto, Y., Kim, C. C., Chang, C.-M., Steere, M., and Fayol, M. (1994). Comparisons of children’s cognitive representation of number: China, France, Japan, Korea, Sweden, and the United States. *Int. J. Behav. Dev.* 17, 401–411. doi: 10.1177/016502549401700301
- Miura, I. T., Okamoto, Y., Vlahovic-Stetic, V., Kim, C. C., and Han, J. H. (1999). Language supports for children’s understanding of numerical fractions: cross-national comparisons. *J. Exp. Child Psychol.* 74, 356–365. doi: 10.1006/jecp.1999.2519
- Moeller, K., Pixner, S., Kaufmann, L., and Nuerk, H.-C. (2009). Children’s early mental number line: logarithmic or rather decomposed linear? *J. Exp. Child Psychol.* 103, 503–515. doi: 10.1016/j.jecp.2009.02.006
- Moeller, K., Pixner, S., Zuber, J., Kaufmann, L., and Nuerk, H.-C. (2011). Early place-value understanding as a precursor for later arithmetic performance – a longitudinal study on numerical development. *Res. Dev. Disabil.* 32, 1837–1851. doi: 10.1016/j.ridd.2011.03.012
- Moeller, K., Shaki, S., Goebel, S. M., and Nuerk, H.-C. (2015). Language influences number processing – a quadrilingual study. *Cognition* 136, 150–155. doi: 10.1016/j.cognition.2014.11.003
- Moura, R., Wood, G., Pinheiro-Chagas, P., Lonnemann, J., Krinzinger, H., Willmes, K., et al. (2013). Transcoding abilities in typical and atypical mathematics achievers. *J. Exp. Child Psychol.* 116, 707–727. doi: 10.1016/j.jecp.2013.07.008
- Ng, S. S. N., and Rao, N. (2010). Chinese number words, culture, and mathematics learning. *Rev. Educ. Res.* 80, 180–206. doi: 10.3102/0034654310364764
- Perry, M., VanderStroep, S. W., and Yu, S. L. (1993). Asking questions in first-grade mathematics classes: potential influences on mathematical thought. *J. Educ. Psychol.* 85, 31–40. doi: 10.1037/0022-0663.85.1.31
- Power, R. J. D., and Dal Martello, M. F. (1990). The dictation of Italian numerals. *Lang. Cognitive Proc.* 5, 237–254. doi: 10.1080/01690969008402106
- Power, R., and Dal Martello, M. F. (1997). From 834 to eighty thirty four: the reading of arabic numerals by seven-year-old children. *Math. Cognit.* 3, 63–85. doi: 10.1080/135467997387489
- Pixner, S., Zuber, J., Hermanova, V., Kaufmann, L., Nuerk, H.-C., and Moeller, K. (2011). One language, two number-word systems and many problems: numerical cognition in the Czech language. *Res. Dev. Disabil.* 32, 2683–2689. doi: 10.1016/j.ridd.2011.06.004
- Salillas, E., and Carreiras, M. (2014). Core number representations are shaped by language. *Cortex* 52, 1–11. doi: 10.1016/j.cortex.2013.12.009
- Seron, X., and Fayol, M. (1994). Number transcoding in children: a functional analysis. *Brit. J. Dev. Psychol.* 12, 281–300. doi: 10.1111/j.2044-835X.1994.tb00635.x
- Song, M.-J., and Ginsburg, H. P. (1987). The development of informal and formal mathematics thinking in Korean and U.S. children. *Child Dev.* 58, 1286–1296.
- Stevenson, H. W., and Lee, S. Y. (1990). Contexts of achievement: a study of American, Chinese, and Japanese children. *Monogr. Soc. Res. Child Dev.* 55, 1–123. doi: 10.2307/1166090
- Stevenson, H. W., Lee, S. Y., and Stigler, J. W. (1987). Mathematics achievement of Chinese, Japanese, and American Children. *Science* 231, 693–699. doi: 10.1126/science.3945803
- Stevenson, H. W., and Nerison-Low, R. (2000). *To Sum it Up: Case Studies of Education in Germany, Japan, and the United States*. Philadelphia, PA: Mid-Atlantic Eisenhower Consortium for Mathematics and Science Education, National Institute on Student Achievement, Curriculum, and Assessment.
- Stevenson, H. W., Stigler, J. W., Lee, S. Y., Lucker, G. W., Kitamura, S., and Hsu, C. C. (1985). Cognitive performance and academic achievement of Japanese, Chinese, and American children. *Child Dev.* 56, 718–734. doi: 10.2307/1129761
- Stigler, J. W., Lee, S. Y., and Stevenson, H. W. (1987). Mathematics classrooms in Japan, Taiwan, and the United States. *Child Dev.* 58, 1272–1285. doi: 10.2307/1130620
- Towse, J. N., and Saxton, M. (1997). Linguistic influences on children’s number concepts: methodological and theoretical considerations. *J. Exp. Child Psychol.* 66, 362–375. doi: 10.1006/jecp.1997.2389
- Vasilyeva, M., Laski, E. V., Ermakova, A., Lai, W. F., Jeong, Y., and Hachigian, A. (2015). Reexamining the language account of cross-national differences in base-10 number representations. *J. Exp. Child Psychol.* 129, 12–25. doi: 10.1016/j.jecp.2014.08.004
- Zuber, J., Pixner, S., Moeller, K., and 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. *J. Exp. Child Psychol.* 102, 60–77. doi: 10.1016/j.jecp.2008.04.003

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Moeller, Zuber, Olsen, Nuerk and Willmes. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# The developmental onset of symbolic approximation: beyond nonsymbolic representations, the language of numbers matters

Iro Xenidou-Dervou<sup>1\*</sup>, Camilla Gilmore<sup>2</sup>, Menno van der Schoot<sup>1</sup> and Ernest C. D. M. van Lieshout<sup>1</sup>

<sup>1</sup> Department of Educational Neuroscience and LEARN! Research Institute for Learning and Education, Faculty of Psychology and Education, VU University Amsterdam, Amsterdam, Netherlands, <sup>2</sup> Mathematics Education Centre, Loughborough University, Loughborough, UK

## OPEN ACCESS

### Edited by:

Ann Dowker,  
University of Oxford, UK

### Reviewed by:

Jennifer M. Zosh,  
Pennsylvania State University,  
Brandywine, USA  
Rebecca Merkley,  
University of Oxford, UK

### \*Correspondence:

Iro Xenidou-Dervou,  
Department of Educational  
Neuroscience and LEARN! Research  
Institute for Learning and Education,  
Faculty of Psychology and Education,  
VU University Amsterdam,  
Van der Boechorststraat 1,  
1081 BT Amsterdam, Netherlands  
i.xenidou-dervou@vu.nl

### Specialty section:

This article was submitted to  
Developmental Psychology,  
a section of the journal  
Frontiers in Psychology

**Received:** 30 October 2014

**Accepted:** 03 April 2015

**Published:** 29 April 2015

### Citation:

Xenidou-Dervou I, Gilmore C, van der  
Schoot M and van Lieshout ECDM  
(2015) The developmental onset of  
symbolic approximation: beyond  
nonsymbolic representations, the  
language of numbers matters.  
*Front. Psychol.* 6:487.  
doi: 10.3389/fpsyg.2015.00487

Symbolic (i.e., with Arabic numerals) approximate arithmetic with large numerosities is an important predictor of mathematics. It was previously evidenced to onset before formal schooling at the kindergarten age (Gilmore et al., 2007) and was assumed to map onto pre-existing nonsymbolic (i.e., abstract magnitudes) representations. With a longitudinal study (Experiment 1), we show, for the first time, that nonsymbolic and symbolic arithmetic demonstrate different developmental trajectories. In contrast to Gilmore et al.'s (2007) findings, Experiment 1 showed that symbolic arithmetic onsets in grade 1, with the start of formal schooling, not earlier. Gilmore et al. (2007) had examined English-speaking children, whereas we assessed a large Dutch-speaking sample. The Dutch language for numbers can be cognitively more demanding, for example, due to the inversion property in numbers above 20. Thus, for instance, the number 48 is named in Dutch “achtenveertig” (eight and forty) instead of “forty eight.” To examine the effect of the language of numbers, we conducted a cross-cultural study with English- and Dutch-speaking children that had similar SES and math achievement skills (Experiment 2). Results demonstrated that Dutch-speaking kindergarteners lagged behind English-speaking children in symbolic arithmetic, not nonsymbolic and demonstrated a working memory overload in symbolic arithmetic, not nonsymbolic. Also, we show for the first time that the ability to name two-digit numbers highly correlates with symbolic approximate arithmetic not nonsymbolic. Our experiments empirically demonstrate that the symbolic number system is modulated more by development and education than the nonsymbolic system. Also, in contrast to the nonsymbolic system, the symbolic system is modulated by language.

**Keywords:** numerical cognition, language, nonsymbolic approximate arithmetic, symbolic approximate arithmetic, kindergarten children, number naming system, symbolic arithmetic development, cross-cultural comparison

## Introduction

Humans and animals seem to be born with an ability to estimate and manipulate abstract magnitudes, namely, nonsymbolic quantities (Flombaum et al., 2005; McCrink and Wynn, 2007;

Cantlon, 2012; Starr et al., 2013; for reviews Dehaene et al., 1998; Feigenson et al., 2004; Dehaene, 2011). This ability has been attributed to the so-called approximate number system (ANS), a cognitive system where nonsymbolic numerosities are assumed to be represented and manipulated (Feigenson et al., 2004; Dehaene, 2011). It is a universal system, which is not affected by cross-cultural differences (Pica et al., 2004). In humans, the precision of the ANS increases with age (Halberda and Feigenson, 2008). But, as humans, we also develop higher-order mathematical abilities, based on the use of arbitrary symbols for representing quantities, for example, Arabic notations. In contrast to abstract nonsymbolic representations, symbolic notations allow us to represent quantities precisely. The ANS is often assumed to be linked with the development of our symbolic mathematical abilities (for a review see Feigenson et al., 2013; but see also the review by De Smedt et al., 2013). Symbolic arithmetic processing with large numerosities in an approximate manner has been demonstrated to onset at the age of 5, before the start of formal schooling (Gilmore et al., 2007) and is often assumed to directly map onto one's readily accessible nonsymbolic representations (Lipton and Spelke, 2005; Gilmore et al., 2007; Mundy and Gilmore, 2009). But is this developmental onset of symbolic arithmetic processing universal? Symbols carry with them their phonological representations, which in turn depend on the language one uses (Carey, 2004; Pica et al., 2004). Thus, even though Arabic symbols are used widely, the way they are named varies significantly across different languages (e.g., Pica et al., 2004; Dehaene, 2011). Early symbolic processing skills have been consistently proven to be significant predictors of math achievement (for a review see De Smedt et al., 2013; see also Göbel et al., 2014b; Lyons et al., 2014), even beyond general processing skills, such as working memory (WM) abilities (Xenidou-Dervou et al., 2013). Therefore, a better understanding of their developmental onset and factors affecting them is rendered necessary. This manuscript investigates, for the first time, the developmental trajectories of nonsymbolic and symbolic arithmetic skills and the roles that development, education and language play in this process.

We often find ourselves in a hurry looking at price tags and making a quick estimation such as: "This package costs 38 euros plus 17 for the extras; that's more than the 50 euros I have with me!" Gilmore et al. (2007) demonstrated that the ability to perform such type of symbolic arithmetic with large numerosities starts at the age of 5, namely before starting primary school instruction. Five years-old children could perform well above chance level on symbolic arithmetic problems, which entailed numbers from 5 to 58. These problems asked for abilities that enable one to give an approximate response, otherwise known as approximation skills (Xenidou-Dervou et al., 2013). Gilmore et al.'s (2007) findings were surprising: this study suggested that children are capable of a form of symbolic arithmetic without needing formal schooling. Of course, the question that rose was how could such young children solve these problems? An explanation was derived from the finding that performance on this type of symbolic arithmetic problems demonstrated exactly the same signature effects as those appearing in corresponding ANS measures, namely in the nonsymbolic versions (Gilmore et al., 2007). It is often assumed that the ANS influences the symbolic

number system (Feigenson et al., 2013) and that symbolic representations directly map onto readily accessible ANS representations (Lipton and Spelke, 2005; Mundy and Gilmore, 2009). The primary signature effect of approximation skills (nonsymbolic or symbolic), is the well-known ratio effect: the more the ratio between two quantities or symbols deviates from 1, the easier it is to compare them (Pica et al., 2004; Barth et al., 2005, 2006; Gilmore et al., 2007, 2010; Xenidou-Dervou et al., 2013, 2014). This is based on the assumption that we perceive numerosities on the basis of a mental number line (Izard and Dehaene, 2008). The further two quantities are from each other, the less their representational overlap on this mental number line and thus the easier it is to compare them. It has also been shown that approximate comparison performance is similar to approximate addition performance (Gilmore et al., 2007).

Since Gilmore et al.'s (2007) study, few have examined the corresponding arithmetic processing skills in such young children. Xenidou-Dervou et al. (2013) assessed kindergarteners' nonsymbolic and symbolic approximation skills in addition and comparison. Using structural equation modeling, Xenidou-Dervou et al. (2013) demonstrated that at the kindergarten stage nonsymbolic approximate addition and comparison load on a single nonsymbolic approximation latent factor, whereas symbolic approximate addition, and comparison load on a distinct factor, that of symbolic approximation. In this study, 5 years-old children performed above chance in all nonsymbolic and symbolic approximation tasks without resorting to known alternative systematic response strategies. They also demonstrated the characteristic ratio effect in all approximation tasks with the exception of one: kindergarteners performance in the symbolic approximate addition task did not demonstrate the ratio effect. Performance in this task was relatively low and close to chance level (56.53%) indicating that the children had difficulties with this task. Furthermore, the authors demonstrated that even though nonsymbolic and symbolic arithmetic processing were related in kindergarten age, they were two distinct abilities (Xenidou-Dervou et al., 2013). These findings provided further proof that symbolic arithmetic, as a linguistically mediated system, does not necessarily map only onto nonsymbolic processing at the kindergarten age (see also Sasanguie et al., 2014).

The fact that kindergarteners performed poorly in symbolic approximate addition in Xenidou-Dervou et al.'s (2013) study and demonstrated no ratio effect contradicted Gilmore et al.'s (2007) findings. Xenidou-Dervou et al. (2013) claimed that this difference might be attributed to task or sample characteristic differences. The symbolic approximate arithmetic tasks used in Gilmore et al. (2007) and Xenidou-Dervou et al.'s (2013) studies differed on certain *task-design characteristics*. The latter entailed a larger range of numerosities (6–70) and the numbers were not read aloud to the children. They merely saw the displayed symbols. These characteristics could have made the task harder and thus might have not captured the onset of the skill in question. Or perhaps the task's design failed to capture the desired ability in general; if that were the case, then one would not expect a ratio effect to appear in grade 1 either. An alternative explanation though could be that the large sample in Xenidou-Dervou et al. (2013) did not have adequate symbolic knowledge to be able

to successfully solve these symbolic arithmetic problems even if they only asked for an approximate response. This would imply that with time and instruction – and thus the gradual automatization of symbols, children’s performance would improve. In other words, the onset of symbolic approximate arithmetic would be expected to take place in grade 1.

Previous studies have shown that precision in nonsymbolic and symbolic magnitude comparison increases with age (Halberda and Feigenson, 2008; Holloway and Ansari, 2009). However, to our knowledge, no previous study has compared the developmental trajectories of nonsymbolic and symbolic *arithmetic* processing. Since symbolic processing necessitates additional cognitive processes related to symbolic knowledge beyond the simple underlying ANS representations, we expected nonsymbolic and symbolic approximate arithmetic to demonstrate different developmental trajectories. As children enter grade 1, they receive formal school instruction and thus acquire symbolic knowledge. Therefore, we hypothesized that symbolic arithmetic would show greater developmental increase compared to the corresponding nonsymbolic arithmetic processing skills. Whereas the characteristic ratio effect in nonsymbolic approximation would be evident across both kindergarten and grade 1, we expected that in symbolic approximate addition it would become significant only after the start of formal schooling, namely in grade 1.

## Experiment 1

### Method

#### Participants

This experiment was part of a large-scale longitudinal project, known as the MathChild project<sup>1</sup>. The project started with 444 kindergarteners ( $M_{\text{age}} = 5.59$  years,  $SD = 0.35$ ) from 25 schools around the Netherlands (for more information, including SES information, please see Participants in Xenidou-Dervou et al., 2013). A year later – in grade 1 – 396 of these children were tested again on the tasks presented in this study. Dropouts were primarily due to changing schools. All dropouts were excluded from the analyses. In grade 1 ( $M_{\text{age}} = 6.50$  years,  $SD = 0.32$ ), the sample consisted of 221 boys and 175 girls. All children spoke Dutch and 95.96% of them had Dutch nationality. Legal guardians’ written consents were received for all children.

#### Procedure

All children were tested individually in quiet settings within the school facilities by trained experimenters, who used a detailed protocol with written instructions. The data reported in this study regard a subset of tasks from the MathChild project. At both time points (kindergarten and grade 1), testing started in November and ended in January of the given academic year. In grade 1, testing included two sessions. The tasks reported in the present study were part of the second session. The order of presentation of the tasks was controlled for by alternating the order of the tasks. Children received small tokens after each session for

encouragement. Kindergarten data have been previously reported in Xenidou-Dervou et al. (2013).

### Materials

Tasks used were computerized and presented in E-prime version 1.2 (Psychological Software Tools, Pittsburgh, PA, USA) with HP Probook 6550b type laptops.

#### *Nonsymbolic approximate addition*

Children saw an image of a girl (Sarah) and a boy (Peter) on the far top left side and right side of the screen correspondingly. A trial entailed the following sequence of steps (see **Figure 1A**): (1) Sarah got an amount of blue dots, (2) These were covered up by a gray box, (3) Then she got some more blue dots, (4) These were now all behind the gray box, (5) Lastly, Peter got some red dots. The question they had to answer was: “Who got more dots?” Participants were instructed to press the blue response box in front of them, if they thought Sarah received more dots, or the red response box, if they thought Peter received more dots. Each animated event lasted 1300 ms and between them there was a 1200 ms interval. Children were instructed to respond as correctly and as fast as possible. Once the red dots appeared on the screen, the children had a maximum of 7000 ms to respond. If they did not respond on time, the trial was automatically coded as incorrect. The fast interchange of events and response process prevented children from counting the dots. Between trials, there was an interval of 300 ms.

Numerosities in this task ranged from 6 to 70. The sum of the blue addends differed with the comparison red addend by three ratios with eight trials in each ratio level: 4/7 (easy ratio), 4/6, (middle), and 4/5 (difficult). Similar to previous studies (Barth et al., 2006, 2008; Gilmore et al., 2010; Xenidou-Dervou et al., 2013, 2014), trials were constructed in a manner that allowed the *post hoc* examination of the use of possible alternative systematic response strategies not related to approximate addition, for example, if children only pressed the red or blue button without adding and comparing the addends (see Appendices in Gilmore et al., 2010; Xenidou-Dervou et al., 2013, 2014). Dots were constructed in MATLAB 7.5 R2007b. As in previous studies, to avoid children’s responses relying on the physical features of the dots, we controlled for dot size, total surface area, total contour length, and density (Barth et al., 2006; Gilmore et al., 2010; Xenidou-Dervou et al., 2013, 2014).

In kindergarten, children received six practice trials in order to optimally comprehend the task (see Barth et al., 2005, 2006; Xenidou-Dervou et al., 2013). In grade 1, they received two practice trials to recall the task’s demands. The task included 24 test trials (see Supplementary Material). No feedback was provided during testing aside from occasional verbal encouragement when necessary.

#### *Symbolic approximate addition*

As in previous studies (e.g., Xenidou-Dervou et al., 2013), this task was identical to its nonsymbolic version with the key difference that the dots were now replaced with blue or red boxes displaying the corresponding Arabic notation (see **Figure 1B**). Children were asked to provide an approximate response, namely they were asked to respond as correctly and as fast as possible to

<sup>1</sup><http://vu.mathchild.nl/en/home/>

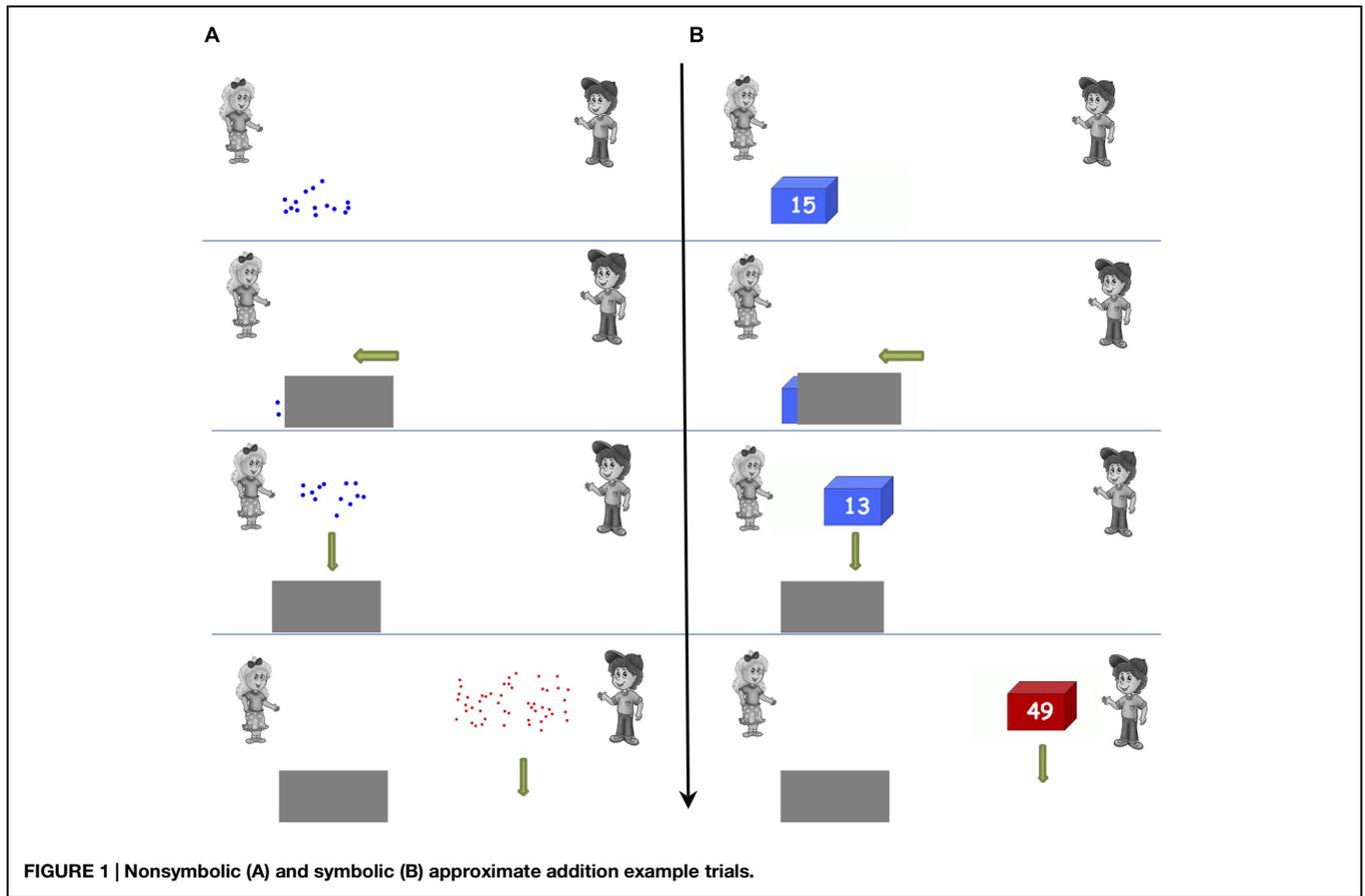


FIGURE 1 | Nonsymbolic (A) and symbolic (B) approximate addition example trials.

the question “Who got more stickers?” The child was asked to estimate, which was more: the sum of the blue number of stickers or the red. The fast interchange of the sequential events and the fact that a response had to be produced within 7000 ms maximum encouraged an approximate response.

**Results**

Children performed above chance level (50%) in all tasks in kindergarten: nonsymbolic addition [ $M = 63.56\%$ ,  $SD = 10.81$ ,  $t(392) = 24.88$ ,  $p < 0.001$ ], symbolic addition [ $M = 57.06\%$ ,  $SD = 11.88$ ,  $t(392) = 11.78$ ,  $p < 0.001$ ] and grade 1: nonsymbolic addition [ $M = 67.76\%$ ,  $SD = 14.19$ ,  $t(395) = 33.12$ ,  $p < 0.001$ ],

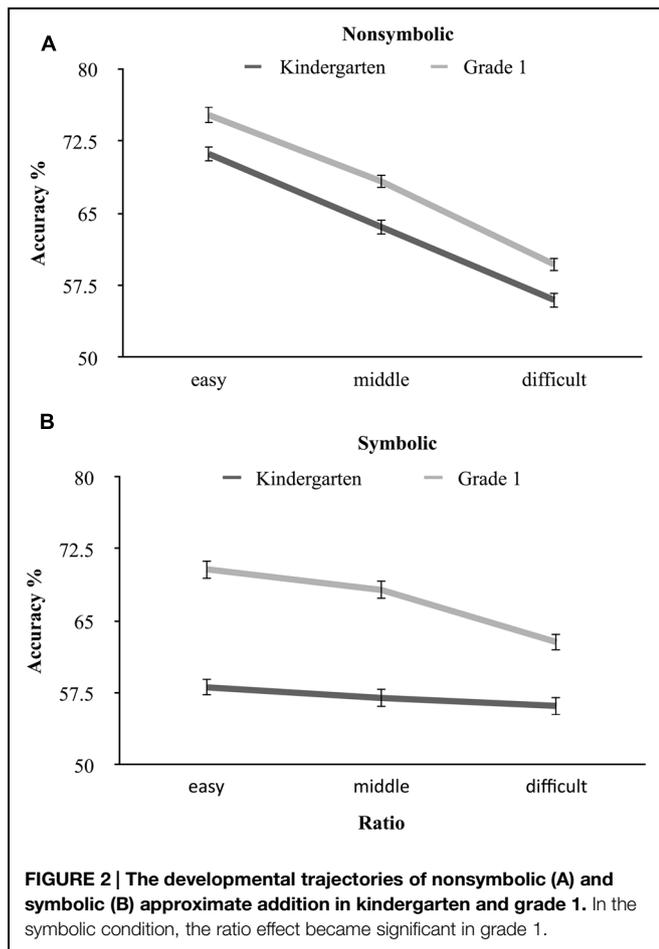
symbolic addition [ $M = 67.07\%$ ,  $SD = 10.81$ ,  $t(395) = 23.93$ ,  $p < 0.001$ ]. Correlations between the assessed measures are presented in **Table 1**.

To compare the developmental trajectories of nonsymbolic and symbolic approximate addition, we conducted a 2 (Task: nonsymbolic and symbolic)  $\times$  3 (Ratio: easy, middle, difficult)  $\times$  2 (Year: kindergarten and grade 1) repeated measures ANOVA. Mauchly’s test indicated that the assumption of sphericity had been violated for Ratio,  $\chi^2(2) = 8.99$ ,  $p = 0.011$ , and the Task by Ratio by Year interaction,  $\chi^2(2) = 12.39$ ,  $p = 0.002$ . Therefore, we corrected the degrees of freedom using Greenhouse–Geisser estimates. Results demonstrated main effects of Task,  $F(1,392) = 37.33$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.09$ , Ratio,  $F(1.96,766.58) = 192.02$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.33$ , Year,  $F(1,392) = 178.72$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.31$ , and the expected significant interaction effect of Task by Ratio by Year,  $F(1.94,760.29) = 3.41$ ,  $p = 0.035$ ,  $\eta_p^2 = 0.01$  (see **Figure 2**). To examine the simple effects two additional analyses were conducted for each task (nonsymbolic and symbolic) separately. For nonsymbolic addition, we found significant main effects of Year,  $F(1,393) = 36.99$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.09$ , and Ratio,  $F(1.97,774.01) = 234.34$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.37$  but not their interaction. For symbolic addition, results showed significant main effects of Year,  $F(1,393) = 196.49$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.33$ , and Ratio,  $F(2,392) = 18.47$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.09$ , but for this

TABLE 1 | Correlations between the nonsymbolic and symbolic arithmetic measures assessed in kindergarten and grade 1.

	1	2	3
(1) Kindergarten nonsymbolic addition			
(2) Kindergarten symbolic addition	0.24*** (393)		
(3) Grade 1 nonsymbolic addition	0.19*** (394)	0.13** (394)	
(4) Grade 1 symbolic addition	0.17*** (394)	0.41*** (394)	0.27*** (396)

Parentheses include the N sample within the specific analysis. \*\* $p \leq 0.01$ ; \*\*\* $p \leq 0.001$ .



task their interaction was also significant,  $F(1.95, 767.78) = 7.29$ ,  $p = 0.001$ ,  $\eta_p^2 = 0.02$ . For this interaction, further simple effect analyses demonstrated that, as expected, in the symbolic condition the ratio effect was only significant in grade 1,  $F(2, 394) = 25.17$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.11$ , and not in kindergarten,  $F(1.95, 764.55) = 1.42$ ,  $p = 0.244$ ,  $\eta_p^2 = 0.00$ . Thus, as hypothesized, nonsymbolic and symbolic approximate arithmetic processing demonstrated different ratio effect developmental trajectories. The ratio effect in symbolic approximate addition became significant in grade 1, **Figure 2B**.

## Discussion

Experiment 1 confirmed our hypothesis that nonsymbolic and symbolic arithmetic processing demonstrate different developmental trajectories. Nonsymbolic acuity increased steadily across time (Halberda and Feigenson, 2008), however, symbolic processing showed a larger increase with the start of formal schooling in grade 1. In other words, symbolic arithmetic processing seemed to be modulated by age and education more than nonsymbolic arithmetic processing. This result also indicated that the symbolic approximate arithmetic task did indeed tap the ability in question: we found a significant ratio effect in symbolic approximate addition in grade 1. So, in this large Dutch sample symbolic approximate arithmetic appeared to onset in grade 1, when

school instruction had started. However, the question remained: why was performance in the symbolic approximate addition task so low at the kindergarten level? Based on Gilmore et al.'s (2007) results, the skill to conduct computations with large symbolic quantities in an approximate manner should start already at the age of 5 years. As described earlier, the difference in results with Gilmore et al. (2007) could still be due to task-design differences. However, there is another, striking difference between the present study and Gilmore et al.'s (2007) study. In Experiment 1, we examined Dutch-speaking children, whereas Gilmore et al. (2007) examined English-speaking children.

There is compelling evidence across interdisciplinary literature demonstrating the importance of the ability to effectively add and compare symbols for children's mathematical achievement (for a review see De Smedt et al., 2013). Given the significant role that symbolic approximation plays in kindergarten math achievement (Xenidou-Dervou et al., 2013), it is imperative to identify the language related factors that play a role in the developmental onset of these skills.

## Experiment 2

So far, the characteristic ratio effect in approximation tasks has been considered universal even when symbols (Arabic numerals) are used. However, the level of transparency of a number naming system has been demonstrated to influence performance even in non-verbal symbolic tasks where the Arabic notation is merely shown, not heard (Nuerk et al., 2005; Helmreich et al., 2011; Göbel et al., 2014a). For example, an essential difference in naming numbers in English versus Dutch (as well as German and other, see Comrie, 2005) is the fact that the latter entail the so-called inversion property. In English, two-digit numbers above 20, such as the number 48, are named in the same order as they are written: first the decades and then the units. In Dutch, however, it is the opposite: first, one names the units and then the decades. So, the number "48" is actually named "eight and forty" (in Dutch: "acht en veertig"). The inversion property has been reported to negatively affect children's symbolic numerical processing. Specifically, Göbel et al. (2014a) demonstrated that it hinders German-speaking (inversion language) second graders' complex two-digit symbolic addition versus their Italian-speaking peers. Furthermore, Helmreich et al. (2011) found German-speaking first graders' number line skills to be less accurate compared to their Italian-speaking peers. Therefore, it could be expected that Dutch-speaking children, similar to German-speaking children would have a disadvantage in their symbolic numerical processing with large numbers due to the demanding Dutch number naming system. Symbolic approximate arithmetic tasks such as those used in Gilmore et al.'s (2007) and the present study entail many two-digit numerosities across their trials and the response on these trials cannot be made by just judging on the basis of the decade of a two-digit number (see Supplementary Material).

Let us consider the cognitive process that could occur when estimating a symbolic number above 20 in English and in

Dutch. In English, the phonological representation of an Arabic two-digit number could involve the following two steps: the child (silently) can vocalize the decades, which he/she then can approximately position on an assumed mental number line. Then, the child can vocalize the units with which he or she fine-tunes approximately the position on the mental number line. In Dutch, the corresponding process appears more demanding. The child first can (silently) vocalize the units but this step would not allow him/her to make an approximate decision on the entire number's position on a mental number line. Instead, this action must be delayed till after the child has vocalized the decades. Meanwhile, the child has to retain the units in his/her WM. In other words, the number naming process in Dutch appears to require more cognitive steps, which will occupy more WM resources. As described earlier, the ratio effect in approximation is assumed to occur because we estimate on the basis of a mental number line where numerosities that are closer to each other have a larger representational overlap and are thus harder to compare. Therefore, the lack of a ratio effect in Dutch kindergarteners' symbolic approximate arithmetic could be due to their demanding number naming system, which would manifest itself as a WM overload. Previous studies have shown that WM is highly related to children's inversion errors when transcoding, namely writing 48 when hearing "forty eight," in German or Czech (Zuber et al., 2009; Pixner et al., 2011). In particular, these studies found that the Central Executive (CE) component of WM, on the basis of to the multicomponent model of WM (Baddeley and Hitch, 1974; Baddeley, 2012), was the most predictive component of inversion-related errors. To our knowledge, the role of WM in symbolic approximate addition in an inversion number naming system such as the Dutch has not been previously addressed.

Cross-cultural studies on numerical skills, thus far, have been conducted with primary school children. Early numeracy skills, however, have been shown to play a role in children's math achievement already from the kindergarten age (e.g., Booth and Siegler, 2006; LeFevre et al., 2010; Mazzocco et al., 2011; Geary et al., 2013; Xenidou-Dervou et al., 2013; Bartelet et al., 2014; Hornung et al., 2014). Furthermore, previous cross-cultural studies did not account for the children's nonsymbolic skills. It could be argued that the groups compared may differ on the basis of their general ability to estimate magnitudes, namely their ANS, not symbolic notations *per se*. We hypothesized that sample differences on the basis of the number naming system children use significantly affects symbolic arithmetic processing beyond their ANS skills. Drawing on the aforementioned assumptions, three clear predictions could be made: (1) Dutch-speaking kindergarteners would have similar ANS skills with matched English-speaking children but would demonstrate a disadvantage in symbolic approximate arithmetic. (2) Dutch-speaking kindergarteners would demonstrate a WM overload in symbolic approximate arithmetic, but not nonsymbolic. (3) The ability to name two-digit numbers would only correlate with symbolic approximate processing, not nonsymbolic. In order to address these hypotheses, we extended our study with a second experiment in

which data was collected from an English-speaking comparison group.

## Method

### Participants

In addition to the existing kindergarten Dutch sample, we tested 54 English-speaking children in the UK ( $M_{\text{age}} = 5.33$  years,  $SD = 0.49$ ; 28 boys). Children, who spoke a second language that entailed the inversion property in their number naming system ( $n = 2$ ) and those with missing data were excluded ( $n = 10$ ). We aimed at having two samples (English-speaking and Dutch-speaking) that had comparable educational and SES backgrounds in order to effectively examine their differences on the basis of language.

With respect to SES, McNeil et al. (2011) have shown that it can influence preschoolers' approximate addition skills. In the present study, SES background was indicated by the parents' level of education. Preliminary analyses in the Dutch sample (used in Experiment 1) had shown that fathers' level of education significantly correlated with their children's symbolic approximate addition ( $r = 0.10$ ,  $p = 0.045$ ). Mothers' level of education did not correlate with the approximation measures. The large Dutch-speaking sample's fathers came from variable SES backgrounds (Xenidou-Dervou et al., 2013). The relatively smaller English-speaking sample, however, consisted of children whose fathers had received higher levels of education. Thus, to control for SES differences across the two samples (UK and NL), children from the NL sample with fathers who had received low educational levels [below HAVO (Dutch educational system)] were excluded from the analyses. The comparison of the two countries' educational systems was based on the official education module comparison developed by the Nuffic (2013), which resulted in seven educational levels. On the basis of these exclusion criteria, the two samples' fathers' SES no longer differed,  $t(47.81) = 0.18$ ,  $p = 0.811$  (NL:  $M = 5.93$ ,  $SD = 0.83$ , UK:  $M = 5.91$ ,  $SD = 0.93$ ).

It was also important that the two samples (UK and NL) had similar educational background. The Dutch kindergarten sample (see Experiment 1) had not received any formal math instruction. Formal instruction in the NL starts in the third year of schooling ("groep 3"). In the UK, however, formal math instruction starts earlier. Therefore, we purposefully assessed younger children in the UK, who had also not yet received formal math instruction. Below the resulting samples from the two countries are described.

The Dutch-speaking sample used in this experiment's analyses consisted of 204 children ( $M_{\text{age}} = 5.58$  years,  $SD = 0.35$ ; 115 boys), 98.04% had Dutch nationality. All children spoke Dutch. According to teacher reports 173 of these children did not speak a second language, for 31 of these children, however, this information was not available as they had moved and changed schools before the time of inquiry. In the Dutch-speaking sample, 92.2% of their fathers and 63.2% of their mothers held an undergraduate or higher academic degree. All the Dutch-speaking children already attended kindergarten ("groep 2" in the Dutch educational system). In this grade in the Netherlands children do not receive structured educational instruction.

The English-speaking sample consisted of 42 children ( $M_{\text{age}} = 5.31$  years,  $SD = 0.53$ ; 23 boys), 97.62% had a UK

nationality. All children spoke English and two of them spoke a non-inversion second language. In this sample, 76.2% of their fathers and 78.6% of their mothers held an undergraduate or higher academic degree. The UK children were tested before the start of the school year during the summer period. At this time the children had only completed 1 year in school. The first year (Reception) is part of the Foundation Stage (age 0–5) during which children learn through play-based activities. In the UK, formal instruction begins during the second year of schooling. As intended, the English-speaking sample was significantly younger compared to the Dutch-speaking sample ( $p = 0.003$ ).

### Procedure

The English-speaking sample was assessed subsequently to the Dutch-speaking sample. Testing took place during the University of Nottingham's Summer Scientist Week<sup>2</sup>. This is an annual research and outreach event during which parents and their children visit the university, play games and take part in studies. SES diversity for this event is highly promoted. Parents/legal guardians provided written consent and SES information. The children were tested in two 20-min sessions. After each session they received tokens to sustain their motivation for participation. For information on the procedure followed in the Dutch sample see Experiment 1. Experimenters in both samples used the same instruction and testing protocol.

### Materials

All the tasks were presented with the same hardware and software as in Experiment 1. The English-speaking sample was assessed on measures that the Dutch sample had been previously tested on (see Xenidou-Dervou et al., 2013). Additionally, the English-speaking sample was also tested on the Naming Large Numbers test.

### Nonsymbolic and symbolic approximate addition

See Experiment 1 (see Materials). The Supplementary Material demonstrates the trials included in this task. It should be noted that in five of these trials (see Supplementary Material, trial numbers: 12, 13, 17, 21, 24) the naming process of their numbers did not differ across English and Dutch on the basis of the inversion property. Since the trials for this task have been stringently constructed based on several control dimensions (see for example Xenidou-Dervou et al., 2013) and due to the comparison with its nonsymbolic counterpart, we opted to keep these five trials. Nevertheless, all trials in the “easy ratio” included two-digit numbers above 20, which are characterized by the inversion property in the Dutch language and not in the English language. We, therefore, expected the difference between the UK and the NL children to be primarily evident in this ratio.

### Exact addition

The exact symbolic addition task (see Jenks et al., 2009; Xenidou-Dervou et al., 2013) assesses children's addition skills in the familiar form of “ $a + b = c$ .” It entailed 15 addition problems, where “a” and “b” were larger than 1 and never equal. The first

10 problems were simple ( $c < 10$ ) and the last five were harder ( $10 < c < 16$ ). The child saw each addition problem on the screen and had to give as correctly and as fast as possible a verbal response for the exact number of the sum. This task demonstrates high levels of internal consistency (Xenidou-Dervou et al., 2013).

### Counting skills

The English and the Dutch version of four subscales from the Early Numeracy Test – Revised (ENT-R, version A) were used to assess children's counting abilities (Van Luit and Van de Rijt, 2009). The subscales assessed (20 items) focused on the child's ability to: (1) use number words (counting forward and backward up to maximum 20); (2) execute structured counting (counting while pointing to objects); (3) conduct resultative counting (counting without pointing to objects); (4), and their general understanding of numbers and how to use the counting system in everyday life.

### Working memory

The English and Dutch versions of two widely known tasks (e.g., Alloway et al., 2004; Xenidou-Dervou et al., 2013) were used to assess children's WM capacity. We had hypothesized that the Dutch number naming system would be phonologically more demanding than the English one. Therefore, we focused on the phonological loop (PL) of the WM construct and its interaction with CE WM resources (Baddeley, 2002; Repovs and Baddeley, 2006).

The *Word Recall Forward* task taps children's PL capacity, namely, the ability to retain phonological information. The child heard a series of recorded high frequency unrelated words and had to repeat them in the same order. After four correct recalls, the child was automatically advanced to the next level that entailed one extra word. A response was registered as correct if the child recalled the word(s) correctly and in the same order as heard. The task would discontinue after three incorrect responses within one level of difficulty.

The *Word Recall Backward* task taps children's CE capacity, specifically the ability to control, regulate and manipulate phonological information. The task's characteristics were identical to the Word Recall Forward task, only now the child was asked to recall the words he/she heard backwards. This task started with a string of two words.

### Naming large numbers test

This test assessed children's ability to name numbers above 20. The children saw a number on the screen, which remained until they gave a verbal response. They were asked to name each number as accurately and quickly as possible. The experimenter pressed a button the moment the child responded, which registered their response time (RT). Nine numbers were used, which are included within the trials of the symbolic approximate arithmetic task and involve the inversion property in the Dutch number naming system but not in the English: 25, 36, 52, 21, 49, 67, 48, 24, and 63 (see Supplementary Material). The order of presentation of the numbers was randomized.

<sup>2</sup>www.summerscientist.org

## Results

### Descriptive Statistics

**Table 2** presents the two groups’ descriptive statistics on the control measures. ANOVAs were conducted to compare performance across the two samples. As expected, they had similar simple addition (in the form of “a + b = c”) and counting skills. However, the Dutch-speaking children had higher WM skills, as they were significantly older than the English-speaking children. We, therefore, controlled for PL and CE WM skills within our subsequent analyses.

### Approximate Addition Comparisons

To examine our first and second hypotheses, we conducted a 3 (Ratio: easy, middle, difficult) × 2 (Country: NL and UK) × 2 (Task: nonsymbolic and symbolic) repeated measures ANCOVA with PL and CE performance as centered covariates (see Thomas et al., 2009). Since the sample sizes across the groups were unequal, Type III sum of squares were used (Maxwell and Delaney, 2004). Box’s *M*-test of equality of covariance matrices in all analyses were not significant. As expected, we found a significant Task by Ratio by Country by CE interaction effect,  $F(2,239) = 4.89, p = 0.008, \eta_p^2 = 0.04$  (see **Figure 3**). In accordance to our hypothesis, CE WM resources appeared to modify the interaction between Task, Country and Ratio. To clarify this 4-way interaction, simple effect analyses were conducted within each task (nonsymbolic and symbolic). For nonsymbolic approximate addition, only the expected main ratio effect was found,  $F(1.89,452.92) = 49.81, p < 0.001, \eta_p^2 = 0.17$ . For the symbolic condition, results demonstrated: a main effect of Ratio,  $F(1.92,460.20) = 6.21, p = 0.003, \eta_p^2 = 0.03$ , a Ratio by Country interaction,  $F(2,239) = 4.73, p = 0.010, \eta_p^2 = 0.04$ , and as expected a Ratio by Country by CE interaction,  $F(2,239) = 5.37, p = 0.005, \eta_p^2 = 0.04$ . Therefore, as hypothesized, the two groups did not differ on the basis of their nonsymbolic approximate skills but only on their symbolic approximate addition performance. Pairwise comparisons indicated that the English-speaking children performed better on the easy ratio of the symbolic approximate addition task ( $p = 0.008$ ), where all trials included an “inversion number.”

To identify the role of the CE component of WM in this interaction, regression equations were constructed with unstandardized regression coefficients on the basis of the parameter estimates derived from the ANCOVA:

$$Y_{\text{easy ratio}} = 66.338 + 4.076X_{\text{CE}} - 8.177X_{\text{country}} - 2.939X_{\text{CE}}X_{\text{country}}$$

$$Y_{\text{middle ratio}} = 56.114 - 0.641X_{\text{CE}} + 0.931X_{\text{country}} + 2.494X_{\text{CE}}X_{\text{country}}$$

$$Y_{\text{difficult ratio}} = 55.186 + 2.575X_{\text{CE}} + 1.721X_{\text{country}} - 1.746X_{\text{CE}}X_{\text{country}}$$

We computed the *Y* values (% symbolic approximate addition performance in each ratio) for 1 *SD* (1.9779) above and below (−1.9711) the mean (0) of the centered CE. In the formulas,  $X_{\text{country}}$  is a dummy variable with the values 0 (UK) and 1 (NL). As depicted in **Figure 4**, for the English-speaking sample, as expected, one notices that with the hypothetical high or low CE value, there are pronounced fluctuations in the ratio effect. Comparing the UK children’s performance with the hypothetical high CE in **Figure 4A** and their performance with the centered (0) CE in **Figure 3B**, the regression equations suggest that the higher their WM capacity, the better their performance was; particularly on the easy ratio of the symbolic task. In the Dutch-speaking sample, however, the ratio effect line remains almost flat no matter the changes in CE values: see **Figures 3B** and **4A,B**. In other words, we see that for the Dutch-speaking children, changes in CE performance do not lead to fluctuations in ratio performance, demonstrating the hypothesized WM overload. Extra CE capacity did not appear to help the Dutch-speaking children; contrary to the English-speaking children it did not appear to facilitate their symbolic approximate addition due to the inversion effect.

### Naming Two-Digit Numbers

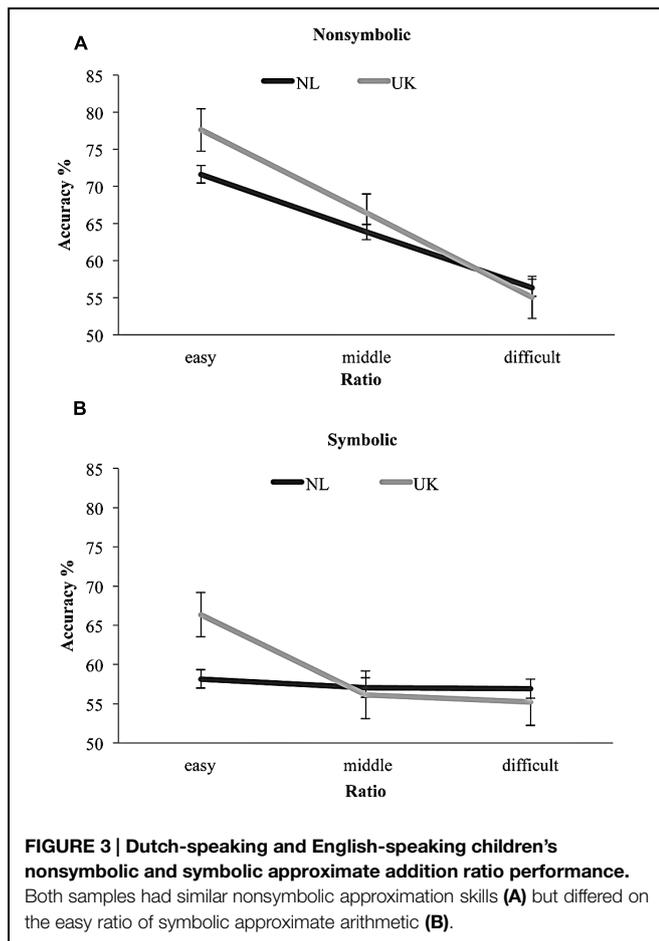
To examine our third hypothesis, we had administered to the whole English-speaking sample ( $n = 52$ ) the “Naming Large Numbers Test.” Results showed that nonsymbolic and symbolic approximate arithmetic correlated significantly  $r = 0.38, p = 0.005$ , but, as expected, the ability to name numbers above 20 correlated highly only with symbolic arithmetic  $r = 0.50, p < 0.001$  and not nonsymbolic  $r = 0.02, p = 0.908$ . Steiger’s *Z*-test (Hoerger, 2013) indicated that these correlation coefficients between the ability to name large numbers and the nonsymbolic and symbolic arithmetic task differed significantly  $Z_H = 3.24, p = 0.001$ .

Accumulatively, our results indicated that number naming characteristics, such as the inversion property entailed in the Dutch number naming system could affect the onset of children’s symbolic approximate arithmetic. We demonstrated that English-speaking children perform better even at a younger age. But can Dutch children even name numbers above 20 at 5 years of age? To answer this question we administered the “Naming Large Numbers Test” to a new Dutch-speaking sample (114 children; 65 boys,  $M_{\text{age}} = 5.4$  years,  $SD = 0.40$ ) matched with the English-speaking one on age ( $p = 0.30$ ). The English-speaking sample could name correctly significantly more two-digit numbers,  $F(1,167) = 7.70, p = 0.006, M_{\text{UK}} (SD) = 5.63 (2.52); M_{\text{NL}} (SD) = 4.34 (2.41)$ , and faster (ms),  $F(1,154) = 135.31, p < 0.001, M_{\text{UK}} (SD) = 2154.29 (1545.11); M_{\text{NL}} (SD) = 10035.28 (4428.58)$ ,

**TABLE 2 | Descriptives and comparisons a cross the two groups.**

Tasks	Country	<i>M</i>	<i>SD</i>	Comparisons
Exact addition	NL	6.22	4.55	ns
	UK	5.12	5.01	
Early numeracy	NL	11.42	4.18	ns
	UK	10.26	4.37	
Phonological loop (PL)	NL	13.29	2.55	$p < 0.001$
	UK	11.64	3.54	
Central executive (CE)	NL	4.65	1.88	$p = 0.001$
	UK	3.55	2.19	

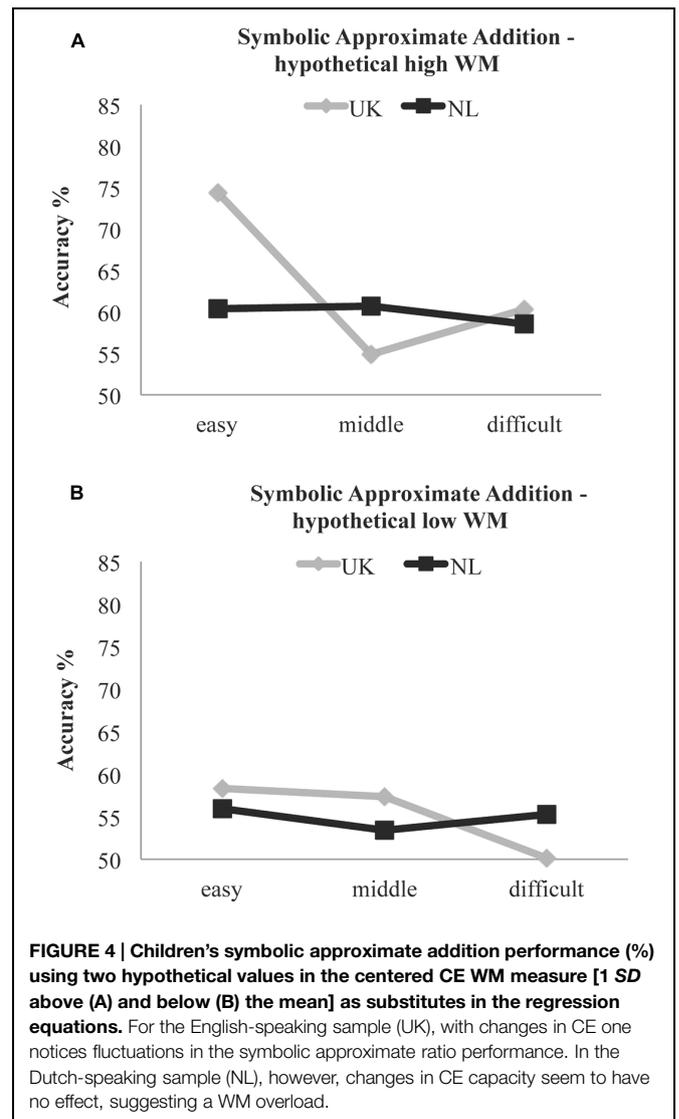
NL, Netherlands; UK, United Kingdom.



than their Dutch-speaking peers. These results showed that at 5 years of age Dutch-speaking children are able to name correctly almost half of the presented two-digit numbers but are worse compared to their English-speaking peers.

## Discussion

In this experiment, we compared English-speaking and Dutch-speaking children's symbolic approximate arithmetic performance controlling for their nonsymbolic approximate arithmetic, simple exact addition and counting skills, as well as WM ability. Also, the two samples did not differ with respect to SES background. Results confirmed our hypotheses. We found that language, specifically differences in the transparency of the number naming system such as the inversion property, can affect the developmental onset of symbolic approximate arithmetic performance. Dutch-speaking kindergarteners lagged behind English-speaking children in symbolic approximate addition, despite being older, and indirectly demonstrated a WM overload in the ratio effect of this form of arithmetic. Furthermore, we found that the ability to name two-digit numbers, which involves the inversion property in Dutch, correlates significantly with symbolic approximation and not nonsymbolic. English-speaking children were better in naming two-digit numbers than their Dutch-speaking peers.



Contrary to Gilmore et al. (2007), who had found the characteristic ratio effect in English-speaking kindergarteners' symbolic approximate addition, Xenidou-Dervou et al. (2013) found no ratio effect in Dutch-speaking kindergarteners' symbolic approximate addition. It should be noted that Gilmore et al.'s (2007) study was conducted with small samples ( $n = 20$ ) drawn from a highly educated community, whereas Xenidou-Dervou et al. (2013) assessed the approximation skills in a large sample, which included a variety of SES backgrounds. But a more pronounced sample difference between the two studies was the language used. The Dutch number naming system involves the cognitively demanding inversion property. Symbolic approximate arithmetic trials involve many two-digit numbers, which entail the inversion property. Previous studies have shown that the inversion property hinders older children's mental number line estimation ability (Helmreich et al., 2011) but had not accounted for the children's general ability to estimate abstract quantities. Our results replicated Gilmore et al.'s (2007) findings,

namely English-speaking 5 years-old performed above chance level and demonstrated the characteristic ratio effect in symbolic approximation. Dutch-speaking kindergarteners, who did not differ with the English-speaking children on SES background and math achievement, had similar nonsymbolic approximation skills. However, as expected, the Dutch-speaking kindergarteners lagged behind the English-speaking children in symbolic approximate addition, even though they were older. Specifically, Dutch children performed worse on the easy ratio, where all test trials included a two-digit number above 20 that needs to be inverted in Dutch (see Supplementary Material). The middle and the difficult ratio of the symbolic approximate addition task were difficult for both the Dutch-speaking as well as the English-speaking children (see **Figure 3B**). In the 4:7 ratio, on the other hand, which is the easiest condition, one would expect that children would have more cognitive resources left to use more effective WM strategies. This was evident for the English-speaking children in **Figure 4A**. For the Dutch-speaking children, however, that was apparently not the case. The two-digit numbers, which need to be cognitively inverted, increased the amount of cognitive resources needed and therefore performance for the Dutch-speaking children was lower than the English-speaking children and the use of effective WM strategies was not feasible (**Figure 3B**).

Nonsymbolic (Xenidou-Dervou et al., 2013, 2014) and symbolic approximation (Caviola et al., 2012; Xenidou-Dervou et al., 2013; Cragg and Gilmore, 2014) necessitate WM resources; especially the CE component of WM as defined by the well-known multicomponent model of WM (Baddeley, 1996, 2002). We had hypothesized that the demanding inversion property would affect Dutch children's symbolic approximation, which entails numbers that are characterized by the inversion property (Zuber et al., 2009; Pixner et al., 2011). When one hears "twenty eight" one can first estimate the position of the number "twenty" on one's mental number line and then refine this position with the use of the "eight." However, when saying "acht en twintig" (eight and twenty) in Dutch, no mental action can be taken with the "acht"; this has to be retained in one's WM and recalled later updating the mental estimation of the "twintig." The ratio effect in approximation is assumed to occur because quantities that are closer to each other have a larger representational overlap on an assumed mental number line. Indeed our results verified that the difference between Dutch- and English-speaking children in symbolic approximation – not nonsymbolic – appeared to be modified by CE capacity. Contrary to the English-speaking children, examining changes in the ratio effect of symbolic approximate addition when increasing CE capacity in the Dutch-speaking sample, one notices no differences in their ratio performance. This demonstrated a significant WM load. In other words, the English-speaking children had room for change/improvement when their CE capacity allowed it, whereas Dutch-speaking children did not. The cognitive load induced by the demanding two-digit Dutch-number naming system was too high at this young age, occupying cognitive resources, which would otherwise allow room for improvement in symbolic approximate addition. It should be noted that in this study we focused on the PL component of WM and its interaction with the CE due to the hypothesized WM load derived from the phonological representations of the numbers.

It would be interesting for future studies, however, to examine also the role of the visuospatial component of WM and its interaction with the CE. Furthermore, future studies should verify our findings with more experimental manipulations in order to demonstrate the causal role of WM within this context.

Furthermore, our results demonstrated for the first time, that the ability to name two-digit numbers correlates highly with symbolic approximation and not nonsymbolic. Previous studies have indicated that the inversion property affects symbolic processing even in non-verbal tasks (Helmreich et al., 2011; Göbel et al., 2014a). It seems that the mere presentation of a number symbol activates its phonological representation in arithmetic. When symbolic approximation is being proven to be an important, consistent predictor of children's math achievement (De Smedt et al., 2013; Xenidou-Dervou et al., 2013), we demonstrate that the ability to name large numbers plays an important role in its developmental onset. Dutch kindergarteners are significantly worse in naming such numbers compared to their English-speaking peers.

The approximate addition tasks used in our experiments entailed two-digit numerosities across all their trials (see Supplementary Material). The trial construction level in this task is stringently balanced across ratios, controlling for alternatives to approximate addition strategy usage and continuous quantity variables in the nonsymbolic condition (see Supplementary Material, also Barth et al., 2006; Gilmore et al., 2010; Xenidou-Dervou et al., 2013, 2014). The inversion effect could potentially affect at any point within an arithmetic process, for example, when merely seeing the numbers in the symbolic condition, when adding them or when comparing the sum to the target quantity. Therefore, in Experiment 2 we used again all trials in order to not disturb the controlled balanced nature of the trials and examine the differences in effect on the basis of the ratio performance. In essence, only two trials in the middle ratio and three trials in the difficult ratio included numbers that do not need to be inverted in Dutch (see Supplementary Material); both of these ratios were hard for all children (see **Figure 3B**). However, all test trials in the easy ratio included an "inversion number" and that is precisely where we found the English-speaking children to outperform the Dutch-speaking children. Our findings cumulatively provide a first indication for the negative effect that the inversion property can have on the onset of symbolic arithmetic. However, future studies should design more rigorous experiments (e.g., Göbel et al., 2014a) targeting specifically the inversion effect on symbolic approximation.

## Conclusion

Cumulatively, findings from both experiments present a clear picture about the importance of education and the language of numbers in developing symbolic arithmetic. Contrary to Gilmore et al. (2007), the present study's results demonstrate that symbolic arithmetic *does* need instruction; it needs instruction of numbers. We showed that development and education modulate symbolic arithmetic more than the ANS. Furthermore, we demonstrated that in contrast to the ANS; symbolic processing is

modulated by language. In Experiment 1, testing a large Dutch sample, we showed that nonsymbolic and symbolic approximate addition have distinct developmental trajectories, with the latter demonstrating significant growth after the start of formal schooling (primary school). In the Dutch-speaking population, symbolic approximate arithmetic onsets in grade 1, not earlier. In Experiment 2, we saw that for English-speaking children, this ability can start earlier. That is because the Dutch number naming system is cognitively more demanding: it involves the inversion property. Our findings demonstrated that Dutch-speaking kindergarteners: (1) Lagged behind English-speaking children in symbolic arithmetic, not nonsymbolic; (2) Demonstrated a WM overload in symbolic approximate arithmetic; not nonsymbolic, and (3) Were significantly worse in naming large numbers compared to their English-speaking peers. Furthermore, we showed that the ability to name large numbers correlated with symbolic, not nonsymbolic approximation. To our knowledge, this is the first evidence for the effect of the inversion property on the onset of symbolic approximation; a core system for the development of mathematical achievement (De Smedt et al., 2013; Xenidou-Dervou et al., 2013).

From a theoretical perspective, our findings demonstrate that while the ANS may be linked with symbolic numerosity processing (Feigenson et al., 2004, 2013; Libertus et al., 2011; Starr et al., 2013; Xenidou-Dervou et al., 2013; Gilmore et al., 2014), developing solid symbolic processing skills goes beyond simple ANS representations. The symbolic number system is modulated more by education and development. Also, language plays an essential role in this process to create solid representations for large exact numbers. Given the extensive research that indicates the importance of symbolic processing skills in the development of children's math achievement (De Smedt et al., 2013; Xenidou-Dervou et al., 2013; Lyons et al., 2014), future studies should place more focus on the role that language plays in developing these skills. From an educational perspective, our results suggest that children who speak languages that entail the inversion property in their number naming system, such as Dutch, German,

Arabic, and other (see Comrie, 2005; Göbel et al., 2011), should place more focus in learning and automatizing two-digit numbers since they are cognitively more demanding compared to other – more transparent – number naming systems. For Dutch-speaking children, our findings suggest that it could potentially be useful to start receiving formal school instruction on Arabic numbers already from kindergarten. There is increasing evidence in older children (Göbel et al., 2011, 2014a; Helmreich et al., 2011) and even adults (Nuerk et al., 2005) on the negative effects the inversion property can have on various mathematical abilities. As a striking example of the importance of this issue, one of our Dutch sample's teachers reported that she overheard a child telling another in class while doing arithmetic: "Just say the numbers in English, it's easier." In times when the transfer of knowledge and skills is prominent and international student assessments prevail, improving early educational instruction is of primary importance.

## Acknowledgments

The authors would like to thank all participating children and parents from the Netherlands and the UK. We would also like to thank Sarah Keeble and Sara Humphries, from Loughborough University and Elise Passchier from the VU University of Amsterdam for all their help. This work was supported by the NWO (National Dutch Organization for Scientific Research) under Grant number PROO 411 07 111. CG is funded by a Royal Society Dorothy Hodgkin Fellowship.

## Supplementary Material

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpsyg.2015.00487/abstract>

## References

- Alloway, T. P., Gathercole, S. E., Willis, C., and Adams, A. (2004). A structural analysis of working memory and related cognitive skills in young children. *J. Exp. Child Psychol.* 87, 85–106. doi: 10.1016/j.jecp.2003.10.002
- Baddeley, A. (1996). Exploring the central executive. *Q. J. Exp. Psychol. A* 49, 5–28. doi: 10.1080/713755608
- Baddeley, A. (2012). Working memory: theories, models, and controversies. *Annu. Rev. Psychol.* 63, 1–29. doi: 10.1146/annurev-psych-120710-100422
- Baddeley, A. D. (2002). Is working memory still working? *Eur. Psychol.* 7, 85–97. doi: 10.1027//1016-9040.7.2.85
- Baddeley, A. D., and Hitch, G. J. (1974). "Working memory," in *Recent Advances in Learning and Motivation*, Vol. 8, ed. G. A. Bower (New York: Academic Press), 47–90.
- Bartelet, D., Vaessen, A., Blomert, L., and Ansari, D. (2014). What basic number processing measures in kindergarten explain unique variability in first-grade arithmetic proficiency? *J. Exp. Child Psychol.* 117C, 12–28. doi: 10.1016/j.jecp.2013.08.010
- Barth, H., Beckmann, L., and Spelke, E. S. (2008). Nonsymbolic, approximate arithmetic in children: abstract addition prior to instruction. *Dev. Psychol.* 44, 1466–1477. doi: 10.1037/a0013046
- Barth, H., La Mont, K., Lipton, J., Dehaene, S., Kanwisher, N., and Spelke, E. (2006). Non-symbolic arithmetic in adults and young children. *Cognition* 98, 199–222. doi: 10.1016/j.cognition.2004.09.011
- Barth, H., La Mont, K., Lipton, J., and Spelke, E. S. (2005). Abstract number and arithmetic in preschool children. *Proc. Natl. Acad. Sci. U.S.A.* 102, 14116–14121. doi: 10.1073/pnas.0505512102
- Booth, J. L., and Siegler, R. S. (2006). Developmental and individual differences in pure numerical estimation. *Dev. Psychol.* 42, 189–201. doi: 10.1037/0012-1649.41.6.189
- Cantlon, J. F. (2012). Math, monkeys, and the developing brain. *Proc. Natl. Acad. Sci. U.S.A.* 109(Suppl. 1), 10725–10732. doi: 10.1073/pnas.1201893109
- Carey, S. (2004). Bootstrapping and the origins of concepts. *Daedalus* 133, 59–68. doi: 10.1162/001152604772746701
- Caviola, S., Mammarella, I. C., Cornoldi, C., and Lucangeli, D. (2012). The involvement of working memory in children's exact and approximate mental addition. *J. Exp. Child Psychol.* 112, 141–160. doi: 10.1016/j.jecp.2012.02.005
- Comrie, B. (2005). "Endangered numeral systems," in *Bedrohte Vielfalt: Aspekte des Sprach(en)tods [Endangered Diversity: Aspects of Language Death]*, eds J. Wohlgemuth and T. Dirksmeyer (Berlin: Weißensee Verlag), 203–230.

- Cragg, L., and Gilmore, C. (2014). Skills underlying mathematics: the role of executive function in the development of mathematics proficiency. *Trends Neurosci. Educ.* 3, 63–68. doi: 10.1016/j.tine.2013.12.001
- Dehaene, S. (2011). *The Number Sense: How the Mind Creates Mathematics, Revised and Updated Edition*. New York: Oxford University Press, 352. Available at: <http://www.amazon.com/The-Number-Sense-Creates-Mathematics/dp/0199753873>
- Dehaene, S., Dehaene-Lambertz, G., and Cohen, L. (1998). Abstract representations of numbers in the animal and human brain. *Trends Neurosci.* 21, 355–361. doi: 10.1016/S0166-2236(98)01263-6
- De Smedt, B., Noël, M.-P., Gilmore, C., and Ansari, D. (2013). How do symbolic and non-symbolic numerical magnitude processing relate to individual differences in children's mathematical skills? A review of evidence from brain and behavior. *Trends Neurosci. Educ.* 2, 48–55. doi: 10.1016/j.tine.2013.06.001
- Feigenson, L., Dehaene, S., and Spelke, E. (2004). Core systems of number. *Trends Cogn. Sci.* 8, 307–314. doi: 10.1016/j.tics.2004.05.002
- Feigenson, L., Libertus, M. E., and Halberda, J. (2013). Links between the intuitive sense of number and formal mathematics ability. *Child Dev. Perspect.* 7, 74–79. doi: 10.1111/cdep.12019
- Flombaum, J. I., Junge, J. A., and Hauser, M. D. (2005). Rhesus monkeys (*Macaca mulatta*) spontaneously compute addition operations over large numbers. *Cognition* 97, 315–325. doi: 10.1016/j.cognition.2004.09.004
- Geary, D. C., Hoard, M. K., Nugent, L., and Bailey, D. H. (2013). Adolescents' functional numeracy is predicted by their school entry number system knowledge. *PLoS ONE* 8:e54651. doi: 10.1371/journal.pone.0054651
- Gilmore, C., Attridge, N., De Smedt, B., and Inglis, M. (2014). Measuring the approximate number system in children: exploring the relationships among different tasks. *Learn. Individ. Differ.* 29, 50–58. doi: 10.1016/j.lindif.2013.10.004
- Gilmore, C. K., McCarthy, S. E., and Spelke, E. S. (2007). Symbolic arithmetic knowledge without instruction. *Nature* 447, 589–591. doi: 10.1038/nature05850
- Gilmore, C. K., McCarthy, S. E., and Spelke, E. S. (2010). Non-symbolic arithmetic abilities and mathematics achievement in the first year of formal schooling. *Cognition* 115, 394–406. doi: 10.1016/j.cognition.2010.02.002
- Göbel, S. M., Moeller, K., Pixner, S., Kaufmann, L., and Nuerk, H.-C. (2014a). Language affects symbolic arithmetic in children: the case of number word inversion. *J. Exp. Child Psychol.* 119, 17–25. doi: 10.1016/j.jecp.2013.10.001
- Göbel, S. M., Watson, S. E., Lervag, A., and Hulme, C. (2014b). Children's arithmetic development: it is number knowledge, not the approximate number sense, that counts. *Psychol. Sci.* 25, 789–798. doi: 10.1177/0956797613516471
- Göbel, S. M., Shaki, S., and Fischer, M. H. (2011). The cultural number line: a review of cultural and linguistic influences on the development of number processing. *J. Cross Cult. Psychol.* 42, 543–565. doi: 10.1177/0022022111406251
- Halberda, J., and Feigenson, L. (2008). Developmental change in the acuity of the “Number Sense”: the approximate number system in 3-, 4-, 5-, and 6-year-olds and adults. *Dev. Psychol.* 44, 1457–1465. doi: 10.1037/a0012682
- Helmreich, I., Zuber, J., Pixner, S., Kaufmann, L., Nuerk, H.-C., and Moeller, K. (2011). Language effects on children's nonverbal number line estimations. *J. Cross Cult. Psychol.* 42, 598–613. doi: 10.1177/0022022111406026
- Hoerger, M. (2013). *Z<sub>H</sub>: An Updated Version of Steiger's Z and Web-Based Calculator for Testing the Statistical Significance of the Difference between Dependent Correlations*. Available at: [http://www.psychmike.com/dependent\\_correlations.php](http://www.psychmike.com/dependent_correlations.php)
- Holloway, I. D., and Ansari, D. (2009). Mapping numerical magnitudes onto symbols: the numerical distance effect and individual differences in children's mathematics achievement. *J. Exp. Child Psychol.* 103, 17–29. doi: 10.1016/j.jecp.2008.04.001
- Hornung, C., Schiltz, C., Brunner, M., and Martin, R. (2014). Predicting first-grade mathematics achievement: the contributions of domain-general cognitive abilities, nonverbal number sense, and early number competence. *Front. Psychol.* 5:272. doi: 10.3389/fpsyg.2014.00272
- Izard, V., and Dehaene, S. (2008). Calibrating the mental number line. *Cognition* 106, 1221–1247. doi: 10.1016/j.cognition.2007.06.004
- Jenks, K. M., de Moor, J., and van Lieshout, E. C. D. M. (2009). Arithmetic difficulties in children with cerebral palsy are related to executive function and working memory. *J. Child Psychol. Psychiatry* 50, 824–833. doi: 10.1111/j.1469-7610.2008.02031.x
- LeFevre, J.-A., Fast, L., Skwarchuk, S.-L., Smith-Chant, B. L., Bisanz, J., Kamawar, D., et al. (2010). Pathways to mathematics: longitudinal predictors of performance. *Child Dev.* 81, 1753–1767. doi: 10.1111/j.1467-8624.2010.01508.x
- Libertus, M. E., Feigenson, L., and Halberda, J. (2011). Preschool acuity of the approximate number system correlates with school math ability. *Dev. Sci.* 14, 1292–1300. doi: 10.1111/j.1467-7687.2011.01080.x
- Lipton, J. S., and Spelke, E. S. (2005). Preschool children's mapping of number words to nonsymbolic numerosities. *Child Dev.* 76, 978–988. doi: 10.1111/j.1467-8624.2005.00891.x
- Lyons, I. M., Price, G. R., Vaessen, A., Blomert, L., and Ansari, D. (2014). Numerical predictors of arithmetic success in grades 1–6. *Dev. Sci.* 17, 714–726. doi: 10.1111/desc.12152
- Maxwell, S. E., and Delaney, H. D. (2004). *Designing Experiments and Analyzing Data: A Model Comparison Perspective*. Mahwah, NJ: Erlbaum.
- Mazzocco, M. M. M., Feigenson, L., and Halberda, J. (2011). Preschoolers' precision of the approximate number system predicts later school mathematics performance. *PLoS ONE* 6:e23749. doi: 10.1371/journal.pone.0023749
- McCrink, K., and Wynn, K. (2007). Ratio abstraction by 6-month-old infants. *Psychol. Sci.* 18, 740–745. doi: 10.1111/j.1467-9280.2007.01969.x
- McNeil, N. M., Fuhs, M. W., Keultjes, M. C., and Gibson, M. H. (2011). Influences of problem format and SES on preschoolers' understanding of approximate addition. *Cogn. Dev.* 26, 57–71. doi: 10.1016/j.cogdev.2010.08.010
- Mundy, E., and Gilmore, C. K. (2009). Children's mapping between symbolic and nonsymbolic representations of number. *J. Exp. Child Psychol.* 103, 490–502. doi: 10.1016/j.jecp.2009.02.003
- Nuerk, H.-C., Weger, U., and Willmes, K. (2005). Language effects in magnitude comparison: small, but not irrelevant. *Brain Lang.* 92, 262–277. doi: 10.1016/j.bandl.2004.06.107
- Nuffic. (2013). *Landenmodule Verenigd Koninkrijk [Country Module United Kingdom]*, 2 Edn. Den Haag: Nuffic.
- Pica, P., Lemer, C., Izard, V., and Dehaene, S. (2004). Exact and approximate arithmetic in an Amazonian indigene group. *Science* 306, 499–503. doi: 10.1126/science.1102085
- Pixner, S., Zuber, J., Heřmanová, V., Kaufmann, L., Nuerk, H.-C., and Moeller, K. (2011). One language, two number-word systems and many problems: numerical cognition in the Czech language. *Res. Dev. Disabil.* 32, 2683–2689. doi: 10.1016/j.ridd.2011.06.004
- Repovs, G., and Baddeley, A. (2006). The multi-component model of working memory: explorations in experimental cognitive psychology. *Neuroscience* 139, 5–21. doi: 10.1016/j.neuroscience.2005.12.061
- Sasanguie, D., Defever, E., Maertens, B., and Reynvoet, B. (2014). The approximate number system is not predictive for symbolic number processing in kindergarteners. *Q. J. Exp. Psychol.* 67, 37–41. doi: 10.1080/17470218.2013.803581
- Starr, A., Libertus, M. E., and Brannon, E. M. (2013). Number sense in infancy predicts mathematical abilities in childhood. *Proc. Natl. Acad. Sci. U.S.A.* 110, 18116–18120. doi: 10.1073/pnas.1302751110
- Thomas, M. S. C., Ansari, D., Jarrold, C., and Karmiloff-Smith, A. (2009). Using developmental trajectories to understand developmental disorders. *J. Speech Lang. Hear. Res.* 52, 336–358. doi: 10.1044/1092-4388(2009/07-0144)
- Van Luit, J. E. H., and Van de Rijdt, B. A. M. (2009). *Utrechtse Getalbegrip Toets-Revised [Early Numeracy Test-Revised]*. Doetinchem: Graviant.
- Xenidou-Dervou, I., De Smedt, B., van der Schoot, M., and van Lieshout, E. C. D. M. (2013). Individual differences in kindergarten math achievement: the integrative roles of approximation skills and working memory. *Learn. Individ. Differ.* 28, 119–129. doi: 10.1016/j.lindif.2013.09.012
- Xenidou-Dervou, I., van Lieshout, E. C. D. M., and van der Schoot, M. (2014). Working memory in nonsymbolic approximate arithmetic processing: a dual-task study with preschoolers. *Cogn. Sci.* 38, 101–127. doi: 10.1111/cogs.12053
- Zuber, J., Pixner, S., Moeller, K., and 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. *J. Exp. Child Psychol.* 102, 60–77. doi: 10.1016/j.jecp.2008.04.003

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Xenidou-Dervou, Gilmore, van der Schoot and van Lieshout. This is an open-access article distributed under the terms of the

*Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*

# On the limits of language influences on numerical cognition – no inversion effects in three-digit number magnitude processing in adults

Julia Bahnmueller<sup>1,2</sup>, Korbinian Moeller<sup>1,2,3</sup>, Anne Mann<sup>2</sup> and Hans-Christoph Nuerk<sup>1,2,3\*</sup>

<sup>1</sup> Knowledge Media Research Center, Tuebingen, Germany, <sup>2</sup> Department of Psychology, Eberhard Karls University, Tuebingen, Germany, <sup>3</sup> LEAD Graduate School, Eberhard Karls University, Tuebingen, Germany

## OPEN ACCESS

### Edited by:

Yvette Renee Harris,  
Miami University, USA

### Reviewed by:

Vrinda Kalia,  
Miami University, USA  
Pedro Macizo,  
University of Granada, Spain

### \*Correspondence:

Hans-Christoph Nuerk,  
Department of Psychology, Eberhard  
Karls University, Schleichstrasse 4,  
72076 Tuebingen, Germany  
hc.nuerk@uni-tuebingen.de

### Specialty section:

This article was submitted to  
Developmental Psychology,  
a section of the journal  
Frontiers in Psychology

**Received:** 28 November 2014

**Accepted:** 31 July 2015

**Published:** 12 August 2015

### Citation:

Bahnmueller J, Moeller K, Mann A  
and Nuerk H-C (2015) On the limits  
of language influences on numerical  
cognition – no inversion effects  
in three-digit number magnitude  
processing in adults.  
*Front. Psychol.* 6:1216.  
doi: 10.3389/fpsyg.2015.01216

The inversion of number words influences numerical cognition even in seemingly non-verbal tasks, such as Arabic number comparison. However, it is an open question whether inversion of decades and units also influences number processing beyond the two-digit number range. The current study addresses this question by investigating compatibility effects in both German- (a language with inverted) and English-speaking (a language with non-inverted number words) university students (mean age 22 years) in a three-digit number comparison task. We observed reliable hundred-decade as well as hundred-unit compatibility effects for three-digit number comparison. This indicates that, comparable two-digit numbers, three-digit numbers are processed in a parallel decomposed fashion. However, in contrast to previous results on two-digit numbers as well as on children's processing of three-digit numbers, no reliable modulation of these compatibility effects through language was observed in adults. The present data indicate that inversion-related differences in multi-digit number processing are limited. They seem to be restricted to the number range involving those digits being inverted (i.e., tens and units in two-digit numbers) but do not generalize to neighboring digits. Possible reasons for this lack of generalization are discussed.

**Keywords:** multi-digit number comparison, three-digit numbers, compatibility effects, language-moderated effects, developmental changes

## Introduction

Everyday life usually involves processing of multi-digit numbers. Nevertheless, much of the research in numerical cognition has been devoted to single-digit number processing. However, findings from single-digit number processing may not simply be transferred to multi-digit number processing (Nuerk et al., 2011). Indeed, specific processes and representations (e.g., base-10 place-value representation, the carry-process in addition) are exclusive to multi-digit number processing (Nuerk et al., 2015). Importantly, such multi-digit number representations are not only of academic interest but seem of particular relevance for numerical development. Moeller et al. (2011) showed that the mastery of the place-value structure of the Arabic number system in first grade predicted later calculation performance. In the following, we will first describe specificities of multi-digit numbers before discussing language influences on multi-digit number processing essential for the current study.

The majority of studies investigating multi-digit number processing focused on two-digit integer numbers (Nuerk et al., 2011, 2015 for overviews). Even though earlier studies concluded on holistic processing of two-digit numbers as integrated entities (e.g., Dehaene et al., 1990; see also Zhang and Wang, 2005; Ganor-Stern et al., 2009) there is accumulating evidence suggesting two-digit numbers to be processed in a decomposed manner (i.e., separated into tens and units, e.g., Nuerk et al., 2001; Ganor-Stern et al., 2007; Kallai and Tzelgov, 2012; Macizo and Herrera, 2013; Moeller et al., 2013).

However, for numbers beyond the two-digit number range empirical evidence is sparser and in contrast to what has been observed for two-digit numbers, recent studies suggest that higher multi-digit numbers are processed in a combined parallel-sequential manner. In an eye-tracking study, Meyerhoff et al. (2012) found that for four- to six-digit numbers processing of the constituting digits becomes less parallel and more sequentially clustered (see also Poltrock and Schwartz, 1984). On the other hand, Korvorst and Damian (2008) observed that three-digit numbers are primarily processed in parallel but with a left-to-right gradient reflecting the relevance of hundreds, tens, and units. Finally, Mann et al. (2012) showed that, for three-digit numbers, parallel decomposed processing developed later and in a less consistent way as compared to two-digit numbers. Taken together, this indicates that results from two-digit numbers cannot simply be transferred to higher multi-digit numbers.

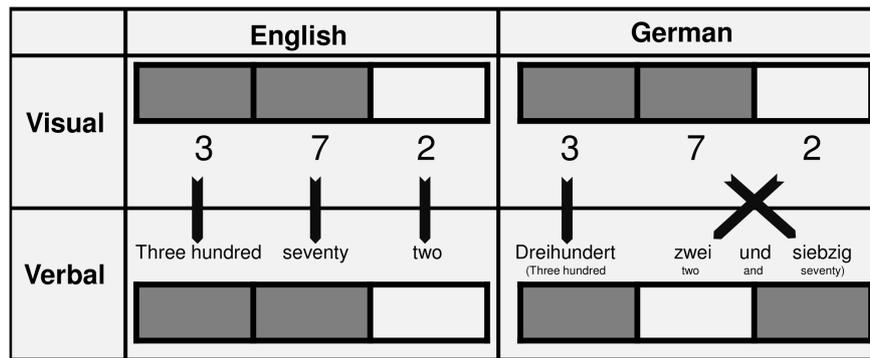
It has long been reported that greater transparency of the number word system facilitates number processing and arithmetic performance at virtually all stages of development (e.g., Miura et al., 1988; Fuson and Kwon, 1991; Miura and Okamoto, 2003; Dowker and Lloyd, 2005; Dowker et al., 2008; Krinzinger et al., 2011). Noteworthy, number word structure influences number processing in verbal but also in other numerical tasks that do not rely on a verbal processing component explicitly. As regards verbal numerical tasks, specific transcoding errors were observed depending on specific characteristics of certain number word systems (e.g., effects of the base-20 system in the French number word system, Seron and Fayol, 1994; see also Colomé, A et al., 2010 for base-20 system effects in Basque). More specifically related to the current study are transcoding errors due to the inversion property of some languages. In German, Dutch, Arabic but also other languages the order of tens and units in number words corresponding to two-digit numbers is inverted as compared to the Arabic digital notation (e.g., 27 ↔ “siebenundzwanzig,” i.e., “seven-and-twenty”). As a consequence, children speaking a language with inversion commit specific inversion related errors in transcoding (i.e., writing down 72 when dictated “seven-and-twenty”). For German-speaking first graders it has been shown that almost 50% of errors committed are inversion related (Zuber et al., 2009). No such errors are reported for languages without inversion (e.g., Italian, cf. Power and Dal Martello, 1990, 1997). In a recent study in Czech, where both an inverted and a regular number word system exist for two-digit number words, revealed the same detrimental effects on transcoding even in a within-participant design in first-grade children (Pixner et al., 2011b; see also Imbo et al., 2014, for cross-linguistic effect within the

same nation). However, number word influences on multi-digit number processing also extend to other numerical tasks such as symbolic magnitude comparison (Nuerk et al., 2005; Pixner et al., 2011a), number line estimation (Helmreich et al., 2011) or mental addition (Göbel et al., 2013).

The effect of interest, with which language influences can be studied in multi-digit number comparison, is the unit-decade compatibility effect. In unit-decade compatible number pairs separate comparisons of tens and units lead to the same decision (e.g., 42\_57;  $4 < 5$  and  $2 < 7$ ) whereas in unit-decade incompatible number pairs separate comparisons of tens and units lead to opposing decisions (47\_62,  $4 < 6$ , but  $7 > 2$ ). Usually, unit-decade compatible pairs are responded to faster and less error prone than incompatible number pairs (e.g., Nuerk et al., 2001; Ganor-Stern et al., 2007, 2009; Moeller et al., 2009; Macizo et al., 2010; Macizo and Herrera, 2013; see Nuerk et al., 2011, 2015, for reviews). Importantly, this effect was shown to be modulated by the numerical distances between corresponding digit positions. Nuerk et al. (2001, see also Nuerk et al., 2004 for children data) found, for example, that the unit-decade compatibility effect was more pronounced when the distance between the unit digits of the two numbers of a number pair was large. This influence of the respective distances indicates that the unit-decade compatibility effect is not an attentional congruity effect or a categorical response conflict but is indeed driven by the separate processing of the numerical magnitudes of the constituent digits of a multi-digit number.

Furthermore, language and namely, inversion of number words, has been shown to influence the compatibility effects in different studies. For two-digit number pairs, it has been shown that the unit-decade compatibility effect is more pronounced for languages with number word inversion (Nuerk et al., 2005, see also Pixner et al., 2011a for children data), at least when numbers are read from left-to-right (Moeller et al., 2015). It has been argued that number word inversion influences the comparison process as the unit digit being named first in the respective number words (erroneously) implies a higher importance and activation of the unit digit, although it is actually irrelevant for the decision. The higher activation of unit digits elevates the compatibility effect, because it is actually a unit interference effect, where the automatic activation of irrelevant unit comparisons cannot be completely suppressed thus hindering or prolonging responses in incompatible trials.

In the current study we were interested in whether and – if so – how three-digit number comparison is influenced by inversion. In languages with non-inverted number words, the order of digits in a three-digit number is the same as the order of constituents of the corresponding number word (e.g., in English: 372 ↔ three-hundred-and-seventy-two). Contrarily, this is not the case in languages with inverted number words (e.g., in German: 372 ↔ three-hundred-two-and-seventy; see **Figure 1**, for illustration). Similar to two-digit numbers, the unit digit is named before the tens digit following the hundreds digit. This may increase interference due to the irrelevant unit digit. However, in contrast to two-digit numbers, the neighborhood of the constituents differs between number words and Arabic numbers. The number word corresponding to the *unit* digit is



**FIGURE 1 | Visual and verbal neighborhood.** If language influences are present, verbally neighboring digits should interfere with the comparison process. In case of a language without inverted number words, the interference

due to the decade digits should be most pronounced. In contrast, for a language with inversion, the interference due to the unit digits should be more pronounced as compared to the decade digits.

the direct neighbor of the number word corresponding to the hundred digit (i.e., three and seven in above example) whereas the direct neighbor of the hundred digit in Arabic digital notation still is the tens digit (i.e., three and two in above example). If linguistic number word structure influences three-digit number processing, interference due to the unit digit might not be restricted to the neighboring Arabic tens digit but might also extend to the verbally neighboring hundred digit. Consequently, interference due to the unit digit should be more pronounced for German-speaking participants whereas for English-speaking participants it was observed that interference due to the decade digit was more pronounced (Korvorst and Damian, 2008; see also **Figure 1** for an exemplary illustration).

Paralleling the unit-decade compatibility effect for two-digit number comparison, both hundred-decade and hundred-unit compatibility can be defined for three-digit number comparison. A three-digit number pair is hundred-decade compatible when separate comparisons of the hundred and the decade digits lead to the same decisions (e.g., 742\_896; hundreds 7 < 8 and decades 4 < 9) and hundred-decade-incompatible when separate comparisons of hundreds and tens lead to opposing decisions (e.g., 362\_517, hundreds 3 < 5, but 6 > 1). Analogously, a three-digit number pair is defined as hundred-unit compatible when separate comparisons of hundred and unit digits yield the same decision (e.g., 742\_896, hundreds 7 < 8 and units 2 < 6) and hundred-unit-incompatible when these separate comparisons lead to opposing decisions (e.g., 537\_692, hundreds 5 < 6, but units 7 > 2). It is important to note that hundred-decade compatibility and hundred-unit compatibility are both attributes of one single number pair, which can be manipulated independently from each other.

So far, there are only few studies investigating language influences in three-digit number comparison tasks. In line with the larger interference due to decade-digits for languages with a non-inverted number word structure, Korvorst and Damian (2008) observed that, for English-speaking adults, the hundred-decade compatibility effect was descriptively more pronounced as compared to the hundred-unit compatibility effect. The authors interpreted this to indicate a left-to-right processing

gradient reflecting partially sequential processing. However, it is important to note that in the original stimulus set of Korvorst and Damian (2008) hundred-decade compatibility was confounded with overall distance: overall distance was larger for hundred-decade compatible number pairs. This may have led to an inflation of the hundred-decade compatibility effect and, therefore, questions the proposed interpretation of a left-to-right processing gradient.

The only direct between-language comparison for three-digit numbers was conducted with children in third grade. Klein et al. (2013) observed that only German-speaking children exhibited a reliable hundred-unit compatibility effect, whereas no such effect was found for Italian-speaking children (a language without inversion). Moreover, the hundred-unit compatibility effect was more pronounced as compared to the hundred-decade compatibility effect for German-speaking third and fourth graders (Mann et al., 2012). This corroborates the argument on more pronounced unit interference when units neighboring hundreds verbally due to the inverted structure of German number words. However, the direct comparisons between the language groups was not significant, so that these differential language influences need to be treated with caution.

It is, however, important to note that with any outcome of above study, children's processing of multi-digit numbers cannot simply be generalized to adults. Children seem to move from a more sequential to a more parallel processing mode for both two-digit (Nuerk et al., 2004; Mann et al., 2011) and three-digit (Mann et al., 2012) numbers. It is well conceivable that language influences children's more sequential processing of three-digit numbers, but does not influence adult's more parallel and more automatic (cf. Kallai and Tzelgov, 2012) processing. Therefore, the question remains whether or not a stronger influence of unit interference in a language with inverted number words may only be a transient developmental phenomenon.

The present study set off to investigate inversion-related language specificities in three-digit number processing in German- (a language with inverted) as well as English-speaking (a language with non-inverted number words) adults. Because hundred-decade compatibility was confounded with overall

distance in the stimulus set of Korvorst and Damian (2008) we created a new better matched stimulus set avoiding such confounds. Nevertheless, in line with results of Korvorst and Damian (2008), we expect to find reliable effects of hundred-decade as well as hundred-unit compatibility for both language groups indicating that three-digit numbers are processed in a parallel-decomposed manner. Yet, for the English-speaking participants, one would expect no differences in the size of hundred-decade and hundred-unit compatibility effects when the larger hundred-decade compatibility effect were indeed driven by the confounded stimulus set of Korvorst and Damian (2008). On the other hand, we had a specific hypothesis regarding the influence of the inversion property of German number words: for our German-speaking participants the interference due to the verbally first named unit-digit should be more pronounced as compared to the English-speaking participants. This interference should result in a relatively larger hundred-unit compatibility effect as compared to the hundred-decade compatibility effect for the German-speaking sample.

The evaluation of the pattern of compatibility effects for the two language groups will provide first empirical evidence on the question of whether the specific influence of the inverted number word structure on numerical cognition generalizes from 2- to 3-digit number processing. In particular, the proposed differential pattern of compatibility effects would corroborate the notion of inversion-related influences in German to persist into adulthood not only for the processing of two- but also of three-digit numbers.

## Materials and Methods

### Participants

Twenty-five native German speakers (three male, four left-handed) and 28 native English speakers (six male, two left-handed) participated in the study. In each group, one participant had to be excluded due to error rates exceeding 10%. Mean age of the resulting samples was  $M = 23.08$  years ( $SD = 6.28$  years) for German- and  $M = 20.11$  years ( $SD = 2.34$  years) for English-speaking participants. Participants were recruited via postings at either the University of Tuebingen or the University of York and received course credits or 5€/4£ for compensation. All participants reported normal or corrected to normal vision. The study was approved by the local ethics committee of the University of York.

### Stimuli and Design

In total, the stimulus set consisted of 640 three-digit number pairs. Numbers containing the same digit more than once (i.e., 545 or 555), multiples of hundred (i.e., 200) and/or multiples of ten (i.e., 420) were not included in the stimulus set. For 320 of these number pairs, all corresponding digits differed from each other. For these critical items, the factors hundred-decade compatibility (HDC) and hundred-unit compatibility (HUC; each compatible vs. incompatible), as well as hundred (HD), decade (DD), and unit distance [UD; for all, small (1–3) vs. large (4–8)] were manipulated orthogonally in a

$2 \times 2 \times 2 \times 2 \times 2$  within-subject design. Problem size (the sum of the two numbers of a number pair) was matched for all resulting 32 conditions and overall as well as decade and unit distance was matched for the respective item conditions. Hundred distance could not be held constant for all conditions, as hundred distance is necessarily smaller for hundred-decade-compatible than hundred-decade-incompatible trials when problem size is held constant across conditions [hundred distances:  $M_{HDC-comp} = 3.7$ ,  $SD_{HDC-comp} = 2.1$ ,  $M_{HDC-incomp} = 4.4$ ,  $SD_{HDC-incomp} = 2.2$ ;  $t(318) = 3.31$ ,  $p < 0.001$ ]. Descriptive characteristics for these 320 critical number pairs are given in the supplementary material.

Additionally, 320 within-hundred number filler pairs were included. Filler items should prevent participants from focusing on decision-relevant hundred digits only. For half of these filler items, hundred digits were held constant (e.g., 475\_421) whereas for the other half hundred and decade digits were identical (e.g., 425\_421).

Number pairs were presented above each other in Arabic notation in white against a black background (font: “Arial,” font size: 24 pt, bold) with a viewing distance of ~50 cm.

### Task and Procedure

In a magnitude comparison task participants had to indicate the larger of two three-digit numbers as fast and accurately as possible. In case the upper number in the display was larger, participants were instructed to press the ‘↑’ key of a standard keyboard with their right index finger. When the lower number was larger, participants had to press the ‘↓’ key with their left index finger. Instructions were given in the respective native language of participants. The two to-be-compared numbers of each pair appeared simultaneously and remained visible until a response key was pressed. Trials were separated by an inter-trial interval of 500 ms. Trial order was randomized separately for each participant. Participants did not receive feedback as to the correctness of their response. Prior to the critical trials, participants performed 10 practice items, which were not part of the stimulus set.

## Results

Only correct responses were considered for analyses [mean error rate was 3.7%,  $SD = 2.1\%$ ; German: 4.0%,  $SD = 2.3\%$ ; English: 3.4%,  $SD = 1.9\%$ ;  $t(49) = 0.86$ ,  $p = 0.392$ ]. All three-digit number pairs with RTs faster than 200 ms were excluded from further analyses. Additionally, all number pairs with RTs deviating more than  $\pm 3$  standard deviation from the individual participant's mean RT were excluded. This procedure led to a total loss of 1.4% of the data [German:  $M = 1.4\%$ ,  $SD = 0.4\%$ ; English:  $M = 1.4\%$ ,  $SD = 0.7\%$ ;  $t(49) = -3.43$ ,  $p = 0.733$ ]. As error rates were very low, analyses focused on RT. Nevertheless, a highly significant positive correlations between error rates and reaction times (German:  $r = 0.788$ ,  $p < 0.001$ ; English:  $r = 0.688$ ,  $p < 0.001$ ) indicated a similar response pattern for errors and RTs disconfirming a speed accuracy trade off. Compatibility and distance effects were evaluated by a  $2 \times 2 \times 2 \times 2 \times 2 \times 2$

**TABLE 1 | Mean and standard deviation for all 32 stimulus categories for the German- and the English-speaking sample.**

Hundred distance	Decade distance	Unit distance		Hundred-decade-compatible		Hundred-decade-incompatible	
				Hundred-unit-compatible	Hundred-unit-incompatible	Hundred-unit-compatible	Hundred-unit-incompatible
				<i>M</i> (SD)	<i>M</i> (SD)	<i>M</i> (SD)	<i>M</i> (SD)
large	large	large	German	680 (141)	729 (186)	693 (152)	700 (166)
			English	724 (158)	754 (162)	738 (155)	758 (138)
large	large	small	German	676 (162)	708 (215)	709 (185)	746 (196)
			English	716 (132)	730 (128)	749 (169)	767 (129)
large	small	large	German	711 (205)	726 (182)	682 (162)	753 (225)
			English	737 (196)	781 (224)	705 (129)	776 (141)
large	small	small	German	698 (155)	684 (158)	706 (172)	722 (187)
			English	739 (148)	724 (109)	773 (162)	777 (165)
small	large	large	German	757 (186)	792 (196)	808 (167)	841 (234)
			English	801 (159)	826 (158)	842 (130)	869 (154)
small	large	small	German	790 (214)	808 (231)	813 (218)	806 (213)
			English	832 (180)	854 (205)	866 (166)	856 (164)
small	small	large	German	680 (141)	807 (186)	755 (218)	839 (172)
			English	803 (138)	879 (203)	761 (113)	876 (141)
small	small	small	German	775 (193)	768 (182)	765 (190)	783 (187)
			English	851 (225)	815 (136)	823 (182)	846 (161)

*M*, mean; *SD*, standard deviation.

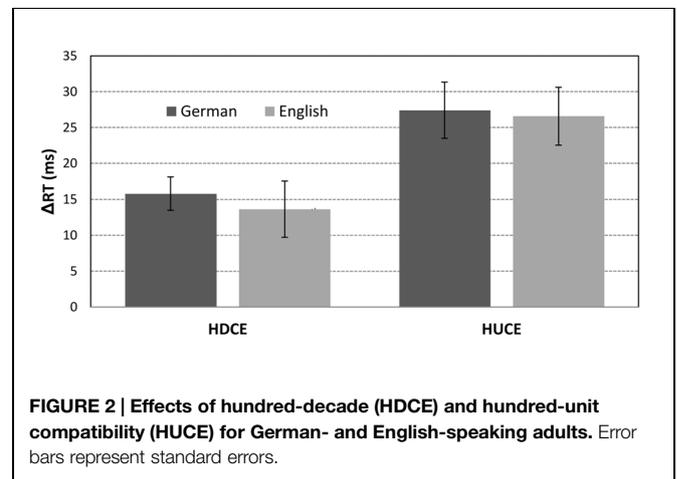
ANOVA discerning the within-subject factors HDC (compatible vs. incompatible), HUC (compatible vs. incompatible), HD (small vs. large), DD (small vs. large), UD (small vs. large) as well as the between-subject factor language (German vs. English). Mean reaction times and standard deviations for all stimulus categories are presented in **Table 1**. In the following we will first report results relevant for possible language differences in compatibility effects, before describing distance effects as well as further modulations of compatibility effects through distances between respective digit positions.

### Compatibility Effects and Language Differences

In line with our hypothesis, reliable main effects of HDC and HUC were observed indicating that three-digit numbers were processed in a parallel and decomposed fashion. Hundred-decade compatible number pairs ( $M = 765$  ms,  $SD = 172$  ms) were on average responded to 14 ms faster than hundred-decade incompatible number pairs [ $M = 779$  ms,  $SD = 165$  ms;  $F(1,49) = 19.65$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.29$ ]. Additionally, response latencies were on average 27 ms shorter for hundred-unit compatible number pairs ( $M = 759$  ms,  $SD = 165$  ms) as compared to hundred-unit incompatible pairs [ $M = 786$  ms,  $SD = 171$  ms;  $F(1,49) = 39.96$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.47$ ].

In contrast to our hypotheses and in contrast to the compatibility pattern observed by Korvorst and Damian (2008), the main effect of HUC was descriptively larger as compared to the effect of HDC for both language groups (see **Figure 2** for an illustration).

In addition to the main effects of HDC and HUC, the interaction between HDC and HUC was significant



[ $F(1,49) = 4.42$ ,  $p = 0.041$ ,  $\eta_p^2 = 0.08$ ]. In contrast to Korvorst and Damian (2008), this interaction indicated that the effect of HUC is larger for hundred-decade *incompatible* as compared to compatible number pairs (33 and 21 ms, respectively).

In contrast to what we expected, neither the interaction between HDC and language [ $F(1,49) = 0.11$ ,  $p = 0.745$ ] nor that between HUC and language [ $F(1,49) = 0.01$ ,  $p < 0.919$ ] nor the three-way interaction of HDC, HUC and language was significant [ $F(1,49) = 0.10$ ,  $p = 0.756$ ]. Overall reaction times did not differ between groups [German;  $M = 750$  ms,  $SD = 196$  ms, English:  $M = 792$  ms,  $SD = 169$  ms;  $F(1,49) = 0.81$ ,  $p = 0.372$ ]. Importantly there wasn't any reliable interaction with the factor language at all. These findings indicate that language did not modulate three-digit number processing. This

interpretation is further corroborated by Bayesian analyses. Using the method proposed by Masson (2011), graded evidence for the null hypothesis (given the obtained data) can be calculated (see Masson, 2011, for a detailed description of the method). With respect to the interaction of both HDC and HUC with language, Bayesian analyses revealed that the probability of the null hypotheses (no differences between language groups) was 0.87 and 0.88, respectively. For the three-way interaction of HDC, HUC, and language the probability was 0.87. Applying the criteria suggested by Masson, probabilities above 0.75 can be considered positive evidence for the null hypothesis.

### Distance Effects and Influences of Digit Distances on Compatibility Effects

A significant main effect of hundred distance was found indexing number pairs with a large HD ( $M = 728$  ms,  $SD = 159$  ms) to be responded to 88 ms faster than pairs with a small HD [ $M = 816$  ms,  $SD = 176$  ms;  $F(1,49) = 534.75$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.92$ ]. This clearly indicates that number magnitude was processed in the task at hand. Moreover, the effect of DD was marginally significant [ $F(1,49) = 3.33$ ,  $p = 0.074$ ,  $\eta_p^2 = 0.06$ ]. In line with the results of Korvorst and Damian (2008) this reflected a tendency toward an inverted DD effect: number pairs with a small DD ( $M = 770$  ms,  $SD = 167$  ms) tended to be processed faster than pairs with a large DD ( $M = 774$  ms,  $SD = 166$  ms). In addition, the interaction of HD and DD was significant [ $F(1,49) = 21.00$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.30$ ] indicating that the reversed DD effect was only present for small hundred distances but not for large HD ( $-15$  and  $7$  ms, respectively).

With respect to the HDC effect, a reliable two-way interaction of HDC and DD was observed [ $F(1,49) = 9.30$ ,  $p = 0.004$ ,  $\eta_p^2 = 0.16$ ]. In line with previous findings on the influence of digit distances on compatibility effects for two-digit numbers (Nuerk et al., 2001) and three-digit numbers (Korvorst and Damian, 2008), this interaction indicated that the HDC effect was more pronounced for larger as compared to smaller DD. Moreover, a reliable interaction of HDC and UD was observed [ $F(1,49) = 4.22$ ,  $p < 0.045$ ,  $\eta_p^2 = 0.08$ ] indicating that the effect of HDC is larger for large UD when compared to small UD. Additionally, three three-way interactions were observed involving HDC. Firstly, the interaction of HDC and DD was further qualified by the three-way interaction of HDC, DD, and HD [ $F(1,49) = 4.49$ ,  $p = 0.039$ ,  $\eta_p^2 = 0.08$ ]. Breaking down this three-way interaction into its constituting two-way interactions revealed that the interaction of HDC and DD was significant for small HD [ $F(1,49) = 11.92$ ,  $p = 0.001$ ,  $\eta_p^2 = 0.19$ ] but not for large HD. Secondly, the interaction of HDC and DD was further qualified by the reliable three-way interaction of HDC, DD, and HUC [ $F(1,49) = 24.16$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.33$ ]. Breaking down this three-way interaction in two-way interactions of HDC and DD for hundred-unit-compatible and -incompatible number pairs showed that the interaction was significant for hundred-unit compatible number pairs [ $F(1,49) = 28.92$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.37$ ] but not for hundred-unit incompatible ones [ $F(1,49) = 0.17$ ,  $p = 0.682$ ]. Lastly, the interaction of HDC, UD, and HD was reliable [ $F(1,49) = 19.48$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.28$ ]. Breaking down this three-way interaction showed that the two-way interaction

of HDC and UD was reliable for large HD [ $F(1,49) = 10.49$ ,  $p = 0.002$ ,  $\eta_p^2 = 0.17$ ] but not for small HD.

For the HUC effect, modulations due to distances between the respective digit positions were observed as well. First, the interaction between HUC and UD was significant [ $F(1,49) = 28.85$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.37$ ] indicating that the effect of HUC was more pronounced for larger as compared to smaller UD. This two-way interaction was further qualified by two three-way interactions including HD and DD, respectively. Breaking down the three-way interaction of HUC, UD, and HD [ $F(1,49) = 8.89$ ,  $p = 0.004$ ,  $\eta_p^2 = 0.15$ ] revealed that the interaction of HUC and UD was reliable for both large [ $F(1,49) = 12.11$ ,  $p = 0.001$ ,  $\eta_p^2 = 0.20$ ] and small HD [ $F(1,49) = 32.32$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.39$ ]. However, while the HUC effect was larger for large UD for both large and small HD, the difference between the HUC effects for large and small UD was more pronounced for small HD. Breaking down the three-way interaction of HUC, UD, and DD [ $F(1,49) = 26.11$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.35$ ] showed that the two-way interaction of HUC and UD was reliable for small DD [ $F(1,49) = 42.44$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.15$ ] but not for large DD.

To further investigate the contribution of inter-digit distances, we ran a regression analysis similar to that conducted by Nuerk et al. (2001) for both language groups separately. Importantly, the results mirrored those of the ANOVA. After checking for collinearity between predictors (e.g., hundred distance was highly correlated with overall distance as well as logarithmic hundred distance), we included the predictors absolute hundred distance, absolute decade distance, absolute unit distance, categorical predictors of HDC and HUC, respectively, continuous predictors of HDC and HUC {e.g., the continuous HDC index for a compatible number pair is positive [732\_896; index: +6 (9-3)] while the continuous HDC index for an incompatible number pair is negative [762\_851; index: -1 (5-6)]}, and problem size (operationalized as the mean of the two to-be-compared numbers). For both language groups, regression analysis was highly predictive [German:  $R = 0.700$ , adj.  $R^2 = 0.482$ ,  $F(5,314) = 60.42$ ,  $p < 0.001$ ; English:  $R = 0.718$ , adj.  $R^2 = 0.507$ ,  $F(5,314) = 66.64$ ,  $p < 0.001$ ] with the same five predictors incorporated in the final model: absolute hundred distance (G:  $b = -0.620$ ; E:  $b = -0.660$ ), HDC categorical (G:  $b = -0.228$ ; E:  $b = -0.219$ ), HUC categorical (G:  $b = -0.196$ ; E:  $b = -0.200$ ), HDC continuous (G:  $b = -0.081$ ; E:  $b = -0.105$ ) and problem size (G:  $b = 0.251$ ; E:  $b = 0.190$ ). Thus, the predictors for both languages were identical. Directly contrasting standardized  $b$ -weights of significant predictors between both languages using the method suggested by Brame et al. (1998) revealed no significant differences for any predictor (all  $Z < 1.1$ , all  $p > 0.27$ ).

In sum, the results of these regression analyses are in line with previous research: the predictor hundred distance was the strongest predictor of RT indicating that participants indeed performed a magnitude comparison task. In addition, results suggest that even for three-digit numbers, compatibility effects are not simply categorical, but are influenced by the magnitudes of the involved digits. This is reflected by the inclusion of the

continuous HDC variable in the final regression model, which did not code compatibility categorically, but defined by magnitude.

## Discussion

The aim of the current study was to investigate whether three-digit number processing in adults is influenced by the specific language property of number word inversion. Although we observed reliable hundred-decade and hundred-unit compatibility effects, these were not modulated by language in the present study. Bayesian analyses substantiated these null effects. In sum, these data argue against a reliable influence of language on three-digit number processing in adults.

This was in contrast to our expectations because such language influences have been found repeatedly for *two-digit numbers* over a variety of tasks and participant groups (e.g., Nuerk et al., 2005; Macizo et al., 2010; Helmreich et al., 2011; Pixner et al., 2011a; Göbel et al., 2013; Imbo et al., 2014). For three-digit number processing, there is first evidence from child data indicating language influences, which were, however, small and not observed consistently (Klein et al., 2013). However, in contrast to this, no language effects were found at all in the present study when evaluating the influence of number word inversion on the pattern of HDC and HUC effects in German- and English-speaking adults using a three-digit number comparison task. This indicates that language influences on multi-digit number processing in children cannot simply be generalized to adults (Nuerk et al., 2004; Mann et al., 2011, 2012).

Nonetheless, comparable and reliable effects of HDC and HUC were observed for both German- and English-speaking adults. This indicated three-digit numbers to be processed in a parallel and decomposed manner in both language groups. Additionally, in both language groups the pattern of compatibility effects differed from the one observed for English-speaking participants previously reported by Korvorst and Damian (2008). In particular, we observed the effect of HUC to be descriptively larger than the effect of HDC contradicting a sequential left-to-right processing gradient in three-digit number processing.

### Lack of Language Differences in Adults

We observed significant HDC and HUC for German and English-speaking adults in ANOVA and regression analyses. In line with previous results for multi-digit numbers (e.g., Nuerk et al., 2001; Ganor-Stern et al., 2007; Macizo et al., 2010; Meyerhoff et al., 2012; Macizo and Herrera, 2013; Moeller et al., 2013) three-digit numbers thus seemed to be processed in a parallel and decomposed manner: the correct decision in number magnitude comparison was not only influenced by the decisive digit (i.e., the hundred digits in the case of between-hundred number pairs) but also by the separate comparisons of decision-irrelevant tens and units. This provides further evidence for the argument that place-value information and the magnitudes of single digits are considered automatically when multi-digit numbers are processed.

Importantly, and in contrast to previous results for two-digit numbers, the present data indicate that three-digit number processing is *not* influenced by number word inversion. The decision-irrelevant tens and units thus exhibited a comparable influence on three-digit number processing for both German- and English-speaking participants. Unlike for tens and units, the position of the hundred digit is not inverted in German number words as compared to its position within the digit string (e.g., 384: *dreihundertvierundachtzig*, literally: *three* hundred four and eighty). Considering this, the present pattern of results indicates that interference due to inverted digits does specifically affect the inverted digits but not those next to the inverted ones. In turn, this might mask potential language differences in multi-digit number processing beyond the two-digit number range. Therefore, the observed compatibility effect pattern indicated that language influences observed for two-digit number processing do not generalize to three-digit number processing. Our results thereby indicate that inversion effects seem to be restricted to the digits being inverted (i.e., tens and units) and do not generalize to the verbally neighboring hundreds.

Please note, however, that the samples investigated in this study comprised only 25 and 28 participants, respectively. Therefore, one might speculate that null effects observed in this study might be attributable to power problems and/or type-1 errors associated with small sample sizes. Yet, this seems unlikely for at least two reasons. First, the observed null effects were substantiated by Bayesian analyses indicating them to be reliable. Second, it needs to be considered that for two-digit numbers, influences of inversion on the compatibility effect have been observed with similar sample sizes (e.g., Nuerk et al., 2005). So, even if statistically significant inversion influences on three-digit number processing might be detected with larger sample sizes, this would provide further evidence for our conclusion that these influences are most likely smaller and/or less reliable than for two-digit number processing in adults.

### Differences between Three-Digit Number Processing in Children and Adults

Investigating an adult sample, the present data did not indicate language to influence three-digit number processing. However, when considering previous evidence from children, data indicate language-specific developmental shifts from more sequential to more parallel decomposed processing of multi-digit numbers (e.g., Klein et al., 2013). As previously observed for two-digit numbers, compatibility effects for three-digit numbers seem to become more pronounced with increasing age. However, for non-inverted languages, the effect of *HUC* was present in (English-speaking) adults (this study and Korvorst and Damian, 2008) but not in (Italian-speaking) children (Klein et al., 2013). For the inverted German language, the effect of *HDC* effect was present in adults but not in children of third and fourth grade. Furthermore, for both German-speaking adults (this study) and children (Mann et al., 2012), the effect of *HUC* effect was larger than the hundred-decade compatibility effect. Thus, there are language influences on the developmental shift from more

sequential toward more parallel decomposed processing of multi-digit numbers with age and experience.

While the shift in processing patterns can be provoked by changes on the visuo-spatial and/or verbal processing level in non-inverted languages (i.e., English, Italian), in the inverted German language, a shift in processing verbal information is more probable. Because the tens did not interfere with the comparison process in German-speaking elementary school children, it can be assumed that, at least for children up to fourth grade, interference caused by verbal number words is more pronounced as compared to interference caused by the directly neighboring digits in symbolic Arabic number notation. Thus, auditory-verbal neighborhood of spoken number word elements seems to be more important at this point than visuo-spatial neighborhood of written digits. This might be due to children's tendency to verbalize what they are cognitively engaged in – in the present case three-digit Arabic numbers – more strongly than adults might do. In line with this, research on so-called private speech indicates that children's use of a more externalized, overt verbal thinking in kindergarten reduces gradually to rather internalized, soundless inner speech over the course of early elementary school (e.g., Kohlberg et al., 1968; Berk, 1992; Winsler and Naglieri, 2003). Although internalizing with age, language-supported processing might still be more pronounced in elementary school children when compared to adults. In turn, visuo-spatial neighborhood seems to become more salient with increasing age, experience, and automaticity in number processing whereas interference due to automatic activation of corresponding number word properties seems to become less salient. Taken together, these findings are in line with the notion of a developmental shift from more *language modulated* sequential processing of three-digit numbers to a more decomposed and parallel processing mode, which gets more independent of language with increasing age and experience.

### Sequential and Parallel Processing and Differential Compatibility Patterns

Combined sequential and parallel processing has been postulated for multi-digit numbers beyond the two digit number range (Korvorst and Damian, 2008; Meyerhoff et al., 2012). For three-digit numbers, Korvorst and Damian (2008) accounted for the larger HDC effect (as compared to the HUC effect) by suggesting a sequential left-to-right processing gradient enhancing the interfering role of the tens. In such sequential processing, tens are assumed to be processed directly following the hundreds and therefore interfere more than the subsequently processed unit digits.

While this explanation is appealing, we were not able to replicate Korvorst and Damian's (2008) results in this respect. Instead, we observed the HUC effect to be descriptively larger than the HDC effect for both the present English- and German-speaking participants. These results are inconsistent with the assumption of a left-to-right processing gradient. As a consequence, the question why differing compatibility effect patterns were found between the present study and that of Korvorst and Damian (2008) is of theoretical importance for our understanding of multi-digit number processing.

To account for this inconsistency, one should first consider differences in the stimulus set. As already described above, overall distance was larger for hundred-decade-compatible as compared to hundred-decade incompatible number pairs in the original stimulus set used by Korvorst and Damian (2008). Therefore, the HDC effect may have been inflated, because overall numerical distance and HDC were confounded. This confound was eliminated in the present stimulus set, which might have led to the smaller hundred-decade-compatibility effect in the present study. Yet, these contrasting results highlight the importance of matching task, stimuli, and procedures when aiming at evaluating multi-digit number processing.

Nevertheless, differences in stimulus characteristics may not be sufficient to explain that – in the present study – the HUC effect was larger than the HDC effect for both language groups. On a first glance, this seems somewhat counterintuitive since hundred and unit digits are visually and conceptually further apart but caused larger inter-digit-interference. A possible explanation for this finding is the effect of lateral masking. Investigated extensively in reading research, this effect describes the interference letters have on the processing of their neighboring letters. When a target letter is flanked by other letters, parafoveal and peripheral vision decreases and, therefore, the probability of correctly identifying the target letter decreases as well (Wolford and Chambers, 1983; Huckauf et al., 1999). Paralleling letter strings or words, one obvious difference between two-digit and three-digit numbers is that only in a three-digit number the decade digit has two neighbors, whereas the hundred and the unit digit only have one. This means less inhibitory influences for the two lateral digits (e.g., hundreds and units) which might in turn add to a descriptively more pronounced HUC effect.

### Conclusion

Taken together, our data suggest that there are limitations to language influences on multi-digit number processing – at least in adults. We observed no influences of number word inversion on three-digit number magnitude processing. This seems counter intuitive because we have seen an increasing number of papers in recent years showing language influences for a wide variety of numerical task, stimulus sets and participant groups. Therefore, the current data constrain these findings: language influences may not be ubiquitous but seem to be specific to stimulus sets, age groups, and probably tasks. Additionally, our data suggest that perceptual determinants of processing multiple elements deserve attention in multi-digit number processing research. On a practical level, future studies might wish to evaluate the development of language differences with age more systematically. Thereby, possible associations with numerical/arithmetic competencies may be investigated (e.g., Göbel et al., 2013), which, in turn, would allow to better understand the influence of inverted number word systems on children's numerical development. On a theoretical level, it would be desirable to better the interplay of parallel and sequential processing of in multi-digit numbers. Therefore, future studies

may use eye-tracking to evaluate online what is actually going on during the comparison process.

## Acknowledgments

JB was supported by the Leibniz-Competition Fund (SAW-2014-IWM-4) providing funding to Elise Klein. KM and H-CN were principal investigators at the LEAD Graduate School [GSC1028], a project of the Excellence Initiative of the German federal and

state governments. We would like to thank Silke Göbel and Carolin Maier for their help in data collection.

## Supplementary Material

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpsyg.2015.01216>

## References

- Berk, L. E. (1992). "Children's private speech: an overview of theory and the status of research," in *Private Speech: From Social Interaction to Selfregulation*, eds R. M. Diaz and L. E. Berk (Hillsdale, NJ: Erlbaum), 17–53.
- Brame, R., Paternoster, R., Mazerolle, P., and Piquero, A. (1998). Testing for the equality of maximum-likelihood regression coefficients between two independent equations. *J. Quant. Criminol.* 14, 245–261. doi: 10.1023/A:1023030312801
- Colomé, À., Laka, I., and Sebastián-Gallés, N. (2010). Language effects in addition: how you say it counts. *Q. J. Exp. Psychol.* 63, 965–983. doi: 10.1080/17470210903134377
- Dehaene, S., Dupoux, E., and Mehler, J. (1990). Is numerical comparison digital? Analogical and symbolic effects in two-digit number comparison. *J. Exp. Psychol. Hum. Percept. Perform.* 16, 626–641. doi: 10.1037/0096-1523.16.3.626
- Dowker, A., Bala, S., and Lloyd, D. (2008). Linguistic influences on mathematical development: how important is the transparency of the counting system?. *Philos. Psychol.* 21, 523–538. doi: 10.1080/09515080802285511
- Dowker, A., and Lloyd, D. (2005). "Linguistic influences on numeracy," in *Mathematics in the Primary School*, eds D. V. Jones, A. Dowker, and D. Lloyd (Oxford: Bangor School of Education), 21–35.
- Fuson, K. C., and Kwon, Y. (1991). "Chinese-based regular and European irregular systems of number words: the disadvantages for English-speaking children," in *Language in Mathematical Education: Research and Practice*, eds K. Durkin and B. Shire (Milton Keynes, PA: Open University Press), 211–226.
- Ganor-Stern, D., Pinhas, M., and Tzelgov, J. (2009). Comparing two-digit numbers: the importance of being presented together. *Q. J. Exp. Psychol.* 62, 444–452. doi: 10.1080/17470210802391631
- Ganor-Stern, D., Tzelgov, J., and Ellenbogen, R. (2007). Automaticity and two-digit numbers. *J. Exp. Psychol. Hum. Percept. Perform.* 33, 483–496. doi: 10.1037/0096-1523.33.2.483
- Göbel, S., Moeller, K., Pixner, S., Kaufmann, L., and Nuerk, H.-C. (2013). Language affects symbolic arithmetic in children: the case of number word inversion. *J. Exp. Child Psychol.* 119, 17–25. doi: 10.1016/j.jecp.2013.10.001
- Helmreich, I., Zuber, J., Pixner, S., Kaufmann, L., Nuerk, H.-C., and Moeller, K. (2011). 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. *J. Cross Cult. Psychol.* 42, 598–613. doi: 10.1177/0022022111406026
- Huckauf, A., Heller, D., and Nazir, T. A. (1999). Lateral masking: limitations of the feature interaction account. *Percept. Psychophys.* 61, 177–189. doi: 10.3758/BF03211958
- Imbo, I., Bulcke, C. V., De Brauwer, J., and Fias, W. (2014). Sixty-four or four-and-sixty? The influence of language and working memory on children's number transcoding. *Front. Psychol.* 5:313. doi: 10.3389/fpsyg.2014.00313
- Kallai, A. Y., and Tzelgov, J. (2012). The place-value of a digit in multi-digit numbers is processed automatically. *J. Exp. Psychol. Learn.* 38, 1221–1233. doi: 10.1037/a0027635
- Klein, E., Bahnmüller, J., Mann, A., Pixner, S., Kaufmann, L., Nuerk, H.-C., et al. (2013). Language influences on numerical development – inversion effects on multi-digit number processing. *Front. Psychol.* 4:480. doi: 10.3389/fpsyg.2013.00480
- Kohlberg, L., Yaeger, J., and Hjertholm, E. (1968). Private speech: four studies and a review of theories. *Child Dev.* 39, 691–736. doi: 10.2307/1126979
- Korvorst, M., and Damian, M. F. (2008). The differential influence of decades and units on multidigit number comparison. *Q. J. Exp. Psychol.* 61, 1250–1264. doi: 10.1080/17470210701503286
- Krinzinger, H., Gregoire, J., Desoete, A., Kaufmann, L., Nuerk, H. C., and Willmes, K. (2011). Differential language effects on numerical skills in second grade. *J. Cross Cult. Psychol.* 42, 614–662. doi: 10.1177/0022022111406252
- Macizo, P., and Herrera, A. (2013). The processing of Arabic numbers is under cognitive control. *Psychol. Res.* 77, 651–658. doi: 10.1007/s00426-012-0456-6
- Macizo, P., Herrera, A., Paolieri, D., and Roman, P. (2010). Is there cross-language modulation when bilinguals process number words? *Appl. Psycholinguist.* 31, 651–669. doi: 10.1017/S0142716410000184
- Mann, A., Moeller, K., Pixner, S., Kaufmann, L., and Nuerk, H.-C. (2011). Attentional strategies in place-value integration. *J. Psychol.* 219, 42–49. doi: 10.1027/2151-2604/a000045
- Mann, A., Moeller, K., Pixner, S., Kaufmann, L., and Nuerk, H.-C. (2012). On the development of Arabic three-digit number processing in primary school children. *J. Exp. Child Psychol.* 113, 594–601. doi: 10.1016/j.jecp.2012.08.002
- Masson, M. E. (2011). A tutorial on a practical Bayesian alternative to null-hypothesis significance testing. *Behav. Res. Methods* 43, 679–690. doi: 10.3758/s13428-010-0049-5
- Meyerhoff, H. S., Moeller, K., Debus, K., and Nuerk, H.-C. (2012). Multi-digit number processing beyond the two-digit number range: a combination of sequential and parallel processes. *Acta Psychol.* 140, 81–90. doi: 10.1016/j.actpsy.2011.11.005
- Miura, I., Kim, C., Chang, C. M., and Okamoto, Y. (1988). Effects of language characteristics on children's cognitive representations of number: cross-cultural comparisons. *Child Dev.* 59, 1445–1450. doi: 10.2307/1130659
- Miura, I., and Okamoto, Y. (2003). "Language supports for mathematics understanding and performance," in *The Development of Arithmetical Concepts and Skills*, eds A. Baroody and A. Dowker (Mahwah, NJ: Lawrence Erlbaum Associates, Inc.), 229–242.
- Moeller, K., Fischer, M. H., Nuerk, H.-C., and Willmes, K. (2009). Sequential or parallel processing of two-digit numbers? Evidence from eye-tracking. *Q. J. Exp. Psychol.* 62, 323–334. doi: 10.1080/17470210801946740
- Moeller, K., Klein, E., Nuerk, H. -C., and Willmes, K. (2013). Magnitude representation in sequential comparison of two-digit numbers is not holistic either. *Cogn. Process.* 14, 51–62. doi: 10.1007/s10339-012-0535-z
- Moeller, K., Pixner, S., Zuber, J., Kaufmann, L., and Nuerk, H.-C. (2011). Early place-value understanding as a precursor for later arithmetic performance—a longitudinal study on numerical development. *Res. Dev. Disabil.* 32, 1837–1851. doi: 10.1016/j.jecp.2010.09.002
- Moeller, K., Shaki, S., Göbel, S. M., and Nuerk, H. C. (2015). Language influences number processing—a quadrilingual study. *Cognition* 136, 150–155. doi: 10.1016/j.cognition.2014.11.003
- Nuerk, H.-C., Kaufmann, L., Zoppoth, S., and Willmes, K. (2004). On the development of the mental number line: more, less, or never holistic with increasing age? *Dev. Psychol.* 40, 1199–1211. doi: 10.1037/0012-1649.40.6.1199
- Nuerk, H.-C., Moeller, K., Klein, E., Willmes, K., and Fischer, M. H. (2011). Extending the mental number line. *Z. Psychol.* 219, 3–22. doi: 10.1027/2151-2604/a000041
- Nuerk, H.-C., Moeller, K., and Willmes, K. (2015). "Multi-digit number processing - overview, conceptual clarifications, and language influences," in *Oxford Handbook of Numerical Cognition*, eds R. Cohen Kadosh and A. Dowker (Oxford: Oxford University Press), 106–139.

- Nuerk, H.-C., Weger, U., and Willmes, K. (2001). Decade breaks in the mental number line? Putting the tens and units back in different bins. *Cognition* 82, B25–B33. doi: 10.1016/S0010-0277(01)00142-1
- Nuerk, H.-C., Weger, U., and Willmes, K. (2005). Language effects in magnitude comparison: small, but not irrelevant. *Brain Lang.* 92, 262–277. doi: 10.1016/j.bandl.2004.06.107
- Pixner, S., Moeller, K., Hermanova, V., Nuerk, H.-C., and Kaufmann, L. (2011a). Whorf reloaded: language effects on nonverbal number processing in first grade—a trilingual study. *J. Exp. Child Psychol.* 108, 371–382. doi: 10.1016/j.jecp.2010.09.002
- Pixner, S., Zuber, J., Hermanova, V., Kaufmann, L., Nuerk, H.-C., and Moeller, K. (2011b). One language, two number-word systems and many problems: numerical cognition in the Czech language. *Res. Dev. Disabil.* 32, 2683–2689. doi: 10.1016/j.ridd.2011.06.004
- Poltrock, S. E., and Schwartz, D. R. (1984). Comparative judgments of multidigit numbers. *J. Exp. Psychol. Learn.* 10, 32–45. doi: 10.1037/0278-7393.10.1.32
- Power, R., and Dal Martello, M. F. (1990). The dictation of Italian numerals. *Lang. Cogn. Process.* 5, 237–254. doi: 10.1080/01690969008402106
- Power, R., and Dal Martello, M. F. (1997). From 834 to eighty thirty four: the reading of arabic numerals by seven-year-old children. *Math. Cogn.* 3, 63–85. doi: 10.1080/135467997387489
- Seron, X., and Fayol, M. (1994). Number transcoding in children: a functional analysis. *Br. J. Dev. Psychol.* 12, 281–300. doi: 10.1111/j.2044-835X.1994.tb00635.x
- Winsler, A., and Naglieri, J. A. (2003). Overt and covert verbal problem-solving strategies: developmental trends in use, awareness, and relations with task performance in children aged 5–7. *Child Dev.* 74, 659–678. doi: 10.1111/1467-8624.00561
- Wolford, G., and Chambers, L. (1983). Lateral masking as a function of spacing. *Percept. Psychophys.* 33, 129–138. doi: 10.3758/BF03202830
- Zhang, J., and Wang, H. (2005). The effect of external representations on numeric tasks. *Q. J. Exp. Psychol.* 58A, 817–838. doi: 10.1080/02724980443000340
- Zuber, J., Pixner, S., Moeller, K., and 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. *J. Exp. Child Psychol.* 102, 60–77. doi: 10.1016/j.jecp.2008.04.003

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Bahnmüller, Moeller, Mann and Nuerk. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# The relation between language and arithmetic in bilinguals: insights from different stages of language acquisition

Amandine Van Rinsveld<sup>1\*</sup>, Martin Brunner<sup>2</sup>, Karin Landerl<sup>3</sup>, Christine Schiltz<sup>1†</sup> and Sonja Ugen<sup>4†</sup>

<sup>1</sup> Education, Culture, Cognition and Society, Institute of Cognitive Science and Assessment, University of Luxembourg, Walferdange, Luxembourg, <sup>2</sup> Berlin-Brandenburg Institute for School Quality, Free University of Berlin, Berlin, Germany, <sup>3</sup> Department of Psychology, University of Graz, Graz, Austria, <sup>4</sup> Luxembourg Center for Educational Testing, University of Luxembourg, Luxembourg, Luxembourg

## OPEN ACCESS

### Edited by:

Hans-Christoph Nuerk,  
University of Tübingen, Germany

### Reviewed by:

Marc Brysbaert,  
Ghent University, Belgium  
Claudia K. Friedrich,  
University of Tübingen, Germany

### \*Correspondence:

Amandine Van Rinsveld,  
Education, Culture,  
Cognition and Society,  
University of Luxembourg,  
Route de Diekirch,  
L-7201 Walferdange, Luxembourg  
amandine.vanrinsveld@uni.lu

†These authors have contributed  
equally to this work.

### Specialty section:

This article was submitted to  
Developmental Psychology, a section  
of the journal *Frontiers in Psychology*

**Received:** 20 October 2014

**Accepted:** 23 February 2015

**Published:** 13 March 2015

### Citation:

Van Rinsveld A, Brunner M, Landerl K,  
Schiltz C and Ugen S (2015) The  
relation between language and  
arithmetic in bilinguals: insights from  
different stages of language  
acquisition. *Front. Psychol.* 6:265.  
doi: 10.3389/fpsyg.2015.00265

Solving arithmetic problems is a cognitive task that heavily relies on language processing. One might thus wonder whether this language-reliance leads to qualitative differences (e.g., greater difficulties, error types, etc.) in arithmetic for bilingual individuals who frequently have to solve arithmetic problems in more than one language. The present study investigated how proficiency in two languages interacts with arithmetic problem solving throughout language acquisition in adolescents and young adults. Additionally, we examined whether the number word structure that is specific to a given language plays a role in number processing over and above bilingual proficiency. We addressed these issues in a German–French educational bilingual setting, where there is a progressive transition from German to French as teaching language. Importantly, German and French number naming structures differ clearly, as two-digit number names follow a unit-ten order in German, but a ten-unit order in French. We implemented a transversal developmental design in which bilingual pupils from grades 7, 8, 10, 11, and young adults were asked to solve simple and complex additions in both languages. The results confirmed that language proficiency is crucial especially for complex addition computation. Simple additions in contrast can be retrieved equally well in both languages after extended language practice. Additional analyses revealed that over and above language proficiency, language-specific number word structures (e.g., unit-ten vs. ten-unit) also induced significant modulations of bilinguals' arithmetic performances. Taken together, these findings support the view of a strong relation between language and arithmetic in bilinguals.

**Keywords:** numbers, language learning, bilingualism, arithmetic, addition

## Introduction

Although every human can manipulate approximate numerical quantities independently from language (Xu and Spelke, 2000), acquiring and mastering symbolic representations of exact quantities critically depends on language and instruction. Amazonian tribes who have restricted or no

number words for quantities larger than five (or even two) impressively illustrate the importance of language for exact quantity representations. While their members can handle and manipulate large numerosities approximately, they are not able to process and represent them exactly (Gordon, 2004; Pica et al., 2004). Formal education enables the acquisition of exact number representations through labeling sets using distinct number names (Fuson et al., 1982). In other words, exact numerical quantities are learned through the use of language (Le Corre and Carey, 2007), and consequently exact number processing remains under the influence of language long after exact number representation acquisition. Recent studies demonstrated that basic processes such as number comparison are performed in slightly different ways depending on task language (Nuerk et al., 2005; Macizo et al., 2010; Van Rinsveld et al., 2012). Yet language plays an especially crucial role in more complex numerical computations, such as arithmetic problem solving. In the present study, we investigated whether and how the progressive acquisition of multiple languages modulates arithmetic problem solving in bilinguals.

### Language and Arithmetic

Several studies provide strong evidence for an involvement of language in exact arithmetic (Spelke and Tsivkin, 2001). Exact calculations, contrary to approximate number processing, is thought to be represented in a specific language-coded format. Neuropsychological studies highlighted that the preservation of language is in fact necessary for arithmetic problem solving, as many authors reported an association between acalculia and aphasia (e.g., Delazer et al., 1999; Basso et al., 2000, 2005, but see Rossor et al., 1995; Cappelletti et al., 2001; Baldo and Dronkers, 2007). In the same way, neuro-imaging studies have shown that exact calculation tasks systematically activate specific language areas, arguing for an exact language-dependent system as opposed to a language-independent approximate system for number representations (Dehaene et al., 1999; Cohen et al., 2000; Stanescu-Cosson et al., 2000; Gruber et al., 2001; Venkatraman et al., 2006, but see Pesenti et al., 2000; Zago and Tzourio-Mazoyer, 2002; Benn et al., 2012).

Language is undoubtedly needed to build exact quantity representations, yet it still has to be clarified for what specific aspect of calculation language plays a crucial role. Heterogeneous solving strategies and processes can be involved in calculation depending on task difficulty (Beishuizen, 1993). Language may consequently affect distinct calculation types differentially. For that matter, it is important to separately examine the specific role played by language in each of the two classically distinguished arithmetic solving strategies. On the one hand, we distinguish simple calculations that are generally composed of one-digit operands (i.e., operands  $< 10$ ). For these problems it is widely accepted that learning and practice lead to a direct retrieval of their solutions from memory, as so-called “arithmetic facts” (Ashcraft, 1992; McCloskey, 1992; Fayol and Thevenot, 2012). However, there is less agreement concerning the storage format of these arithmetic facts: they could be represented in an abstract semantic format (McCloskey et al., 1985), a verbal format (Campbell, 1994; Dehaene and Cohen, 1995) or in a format depending on

individual subject’s preferences (Noel and Seron, 1993). Moreover, the importance of language in arithmetic fact retrieval is modulated by operation-type: simple addition and multiplication, in comparison to subtraction and division, especially rely on verbally coded facts probably because addition and multiplication facts are more often learned and used in their verbal code than subtraction and division (Lemer et al., 2003).

More complex calculations (i.e., operands  $> 10$ ), on the other hand, cannot directly be retrieved from memory but they require mental computations to be solved. These computations mainly rely on working memory resources to execute solving strategies, keep intermediate solutions in memory and update the final solution (Hitch, 1978; Ashcraft, 1995). According to Logie et al. (1994), the phonological loop of Baddeley’s working memory model (Baddeley, 1992) is used in mental calculation to verbally repeat the numbers. Studies using articulatory suppression during complex calculation have shown that phonological mediation occurs, especially when some elements of the problems disappear after a short presentation time (Fürst and Hitch, 2000). Moreover, a study with English-Welsh bilinguals found longer response times and more errors when calculations were performed in Welsh than in English, due to longer number words in the former (Ellis and Hennessey, 1980). Klessinger et al. (2012) revealed that the impact of number word lengths on exact additions was especially prominent in less proficient calculators. Similarly, a neuro-imaging study confirmed the crucial role of working memory in complex calculation and suggested that the working memory components engaged in the (visual or verbal) solving process may depend on individual solving strategies (Delazer et al., 2003). Taken together, these findings suggest that language is important for arithmetic problem solving at different levels because arithmetic facts are potentially represented or retrieved from memory in a verbal format and complex arithmetic solving processes rely at least partially on verbal working memory components.

Language is crucial for exact representation of large numerosities and for exact arithmetic problem solving, at least during the acquisition of these abilities. The different number naming systems used in different languages can modulate numerical performances during acquisition stages of numerical cognition but also in adults, who have long acquired these abilities (Campbell and Xue, 2001; Chen et al., 2009). Specifically, the order of tens and units in two-digit number words is a characteristic of number naming systems that can directly affect arithmetic performances. Brysbaert et al. (1998) showed that additions presented in the format “21 + 4” were solved faster by French-speaking participants (21 is pronounced twenty and one since French number words follow the ten-unit order), whereas the same additions presented in the format “4 + 21” were solved faster for Dutch-speaking participants (21 is pronounced one and twenty since Dutch number words following the unit-ten order). Moreover, Göbel et al. (2014) reported that German-speaking children had a larger carry effect in additions compared to Italian-speaking children. They explained this result by the greater similarity between the Arabic digit notation and the order of tens and units in Italian number words than in German number words. Indeed, Arabic digit notation follows the same order (from left to right) than the

tens-units order of Italian number words (e.g., “24” = twenty-four) but they are inverted in comparison to the unit-ten order of German number words (e.g., “24” = four-and-twenty).

Another difference between number naming systems from different languages is the use of a base-20 structure instead of base-10 structure for two-digit number words (e.g., in French and in Basque). Seron and Fayol (1994) highlighted the specific difficulties encountered by French-speaking children for 70 and 90 number words following this base-20 structure used in France, in comparison to Belgian French-speaking children who use base-10 structure for 70 and 90. The former took longer and made more errors than the latter when writing down Arabic digits in a number dictation task. The base-20 system seems to have an impact not only during development but also later since a study by Colomé et al. (2010) showed that adult Basque speakers are influenced by the base-20 system of their language when they solve addition problems (see also Salillas and Carreiras, 2014). Taken together, the structure of the number words in the language in which numbers are acquired appears to affect arithmetic performances during childhood and some influence on arithmetic computation even persists in adulthood.

## Bilinguals and Arithmetic

Given the critical role of language in arithmetic problem solving, how people using several languages (e.g., bilinguals) calculate is a particularly intriguing question. Many models of language processing in bilinguals support the idea that bilinguals' languages are active in parallel at any time, occasioning mutual interferences between languages (e.g., Kroll et al., 2014). It is generally assumed that interferences of the dominant language during the use of the non-dominant language are more consequential than interferences in the opposite direction, and the relative asymmetry in this mutual influence is a function of bilingual proficiency (Bialystok, 2009; Kroll et al., 2013). Indeed, higher proficiency level in a language lessons potential interferences of other languages on it. From the reports in the literature about bilingual's arithmetic problem solving it appears that several elements concerning the relative mastery of languages (i.e., language proficiency) as well as the structure of the number words in the involved languages directly modulate bilinguals' performances in arithmetic. Relevant data concerning these two aspects are highlighted below.

Early studies in bilingual speakers provided first indications that arithmetic skills is related to language proficiency. They indeed observed systematic advantages in response time and accuracy when bilinguals calculated in their first compared to their second (and less proficiently mastered) language (Marsh and Maki, 1976; McClain and Huang, 1982; Geary et al., 1993). French-Mestre and Vaid (1993) tested addition fact-verification tasks in bilinguals with correct-outcome problems but also false-outcome problems that could be related or unrelated to multiplication facts (i.e.,  $2+3=6$  was a false-outcome addition related to a multiplication fact). The authors observed associative confusion when problems were presented in bilinguals' first language and in Arabic digits but not in bilinguals' second language, so that they argued in favor of automatic arithmetical fact retrieval in the first but not the second language.

More recently, neuro-imaging studies on late Chinese-English bilinguals suggested that the verbal code of the first language is needed to retrieve arithmetic facts when the network of arithmetic facts in second language is not sufficiently developed. Wang et al. (2007) observed that performing complex calculations in first and second languages rely on a common activation network, but with higher activations during calculations in second language. This was interpreted as evidence for extra language processing needs in second language; potentially translation of input from second into first language (Lin et al., 2011). Taken together, these bilingual studies point toward an advantage for both retrieving arithmetic facts and computing complex arithmetic problems in the first language, i.e., the language in which most bilinguals learned to do arithmetic.

However, the bilinguals tested in the aforementioned studies were all late bilinguals or clearly unbalanced bilinguals, so the picture may be a bit different in more balanced bilinguals or bilinguals who did not acquire arithmetic in their first language. Indeed, one study reported that highly proficient bilinguals produced arithmetic facts equally well in their two languages (Campbell and Epp, 2004). Moreover, a study with Filipino-English bilinguals reported better arithmetic fact verification performances in English number word presentation, which was their second language but also the language in which they learned arithmetic at school and which they reported as their preferred language for doing arithmetic (Bernardo, 2001). Furthermore, Salillas and Wicha (2012) provided evidence for strong associative networks between terms and solutions for problems in the language in which participants learned arithmetic, which was not necessarily their first language. Participants seemed to maintain these early-established networks in adulthood, independently of language proficiency. These results were supported by a recent study where bilinguals showed switching costs when they had to retrieve arithmetic facts in their untrained- vs. trained-language (Saalbach et al., 2013). Hence, when bilinguals solve arithmetic problems, the language in which arithmetic was learned might be even more critical than the first language or the language in which they are currently most proficient.

In sum, bilinguals' arithmetic performances can be modulated by language proficiency levels, language of math acquisition and number word structure of the respective spoken languages. However, we are still lacking extensive studies, which investigate the relation between these different factors and arithmetic performances in bilingual participants. Such approaches are nevertheless necessary to understand in detail how language contributes to numerical computations. It is for instance currently unclear whether in highly proficient bilinguals performance levels in arithmetic become equivalent for their two languages or whether they maintain an advantage for retrieving and/or calculating in the language of arithmetic acquisition. It is also not known what increasing language proficiency implies for simple and complex calculations. Finally, it remains to be explored how language-related differences in number word structures affect arithmetic performance in bilinguals (e.g., German units-tens vs. the French tens-units; German base-10 vs. French base-20 for number words between 70 and 99).

## The Present Study

One of the major issues when studying bilinguals is that there often are as many different stories and profiles of languages acquisition as individuals. However, age of acquisition and proficiency levels of languages in bilinguals may drastically influence various ranges of cognitive processes (e.g., Altarriba and Basnight-Brown, 2007). In the present study we took advantage of the unique German-French bilingual school system of Luxembourg in order to address the aforementioned questions concerning to the relation between language and arithmetic by tracking the development of addition solving in bilingual adolescents and young adults at five different stages of bilingual proficiency. Bilingualism is a major attribute of the Luxembourgish educational system, as German and French are both teaching languages. In primary school teaching is held exclusively in German, but during secondary school, teaching language progressively switches to French, so that the pupils become highly proficient both in German and French through their education.

We composed four samples of German-French bilingual pupils at different levels of Luxembourgish secondary school (i.e., grades 7, 8, 10, and 11) and one sample of German-French bilingual young adults (who had also attended secondary school in Luxembourg). All participants thus mastered both German and French. Pupil participants from grades seven to 11 incrementally improved their mastery in German and French, with a relative emphasis on French as this language was becoming their predominant teaching language. The young adults achieved the level of excellence in both French and German. Altogether this yielded a design encompassing five distinct stages of German-French bilingualism.

For empirical research on the interplay of language and arithmetic the bilingual context in Luxembourg is characterized by a double advantage. (a) Firstly, all participants of a given age-class have a similar exposure to each of the two languages, as they are all first taught in German and then in French. This allows composing large samples of bilingual participants that are homogenous in terms of duration and amount of exposure to each language. Moreover, although bilingual, all participants acquired arithmetic in German. (b) Secondly, German-French bilinguals are particularly interesting because German and French languages use inverted number word structures. Two-digit number words follow the units-tens order in German (e.g., “four-and-twenty”) but the tens-units order in French (e.g., “twenty-four” like in English).

The experimental tasks consisted in addition problems that participants had to solve both in German and in French during two separate sessions. Additions were presented in two different formats (i.e., visual presentation of Arabic digits and auditory presentation of number words) and consisting of two difficulty levels (i.e., simple and complex additions). Throughout the entire experiment participants had to give their answers orally in the language of the session. Thus, task language permeated task instructions as well as presentation and solution of the additions in the auditory format, whereas only task instructions and solution production were imbued by task language in the visual format. Based on the literature reviewed above a series of predictions concerning the influence of language on addition solving

in German-French bilinguals could be derived. Moreover, we also formulated detailed proposals on how these language effects might express at different stages of bilingual proficiency.

## Effects of Calculation Complexity on Bilinguals' Arithmetic Solving

In simple additions participants are thought to retrieve the solution from memory (Ashcraft, 1995). Previous studies with bilinguals have shown evidences for early-encoded arithmetic facts in one of the bilinguals' languages (e.g., Frenck-Mestre and Vaid, 1993; Spelke and Tsivkin, 2001; Wang et al., 2007) underlined by format-depending representations (Dehaene and Cohen, 1995). Nevertheless, other studies have highlighted evidences for transferable facts from one language to the other in very proficient bilinguals (e.g., Campbell and Epp, 2004) suggesting a possible representation of numbers independent from any format or language of encoding (McCloskey et al., 1985). Consequently, it can be expected that highly proficient bilinguals (i.e., adults and older adolescent participants of the present study) retrieve addition facts equally well in German and French. Indeed, these participants should be proficient enough in French and/or have been sufficiently exposed to numbers in French to be able to solve the simple additions similarly in French as in German.

Language-related performance differences ought to predominantly arise with complex additions. Compared to simple calculations, more complex arithmetic problems are thought to rely on computational procedures composed of multiple processing steps (e.g., Fayol and Thevenot, 2012), which can be modulated by language proficiency but also by specific number word structures. With respect to task presentation format, language effects should be larger for auditory than visual presentation formats because the operands have to be kept in memory in the former (LeFevre et al., 2001). In line with the prominent role of language, we expected that participants of all proficiency levels solve complex additions better and faster in German than in French. Indeed all participants had acquired German earlier than French and German was also their language of arithmetic acquisition (Bernardo, 2001; Salillas and Wicha, 2012). At the highest bilingual proficiency levels this benefit should be reduced, but we anticipated that it might never be resorbed completely if the early constellation of bilingual proficiency is critical.

## Effects of Number Word Structures on Bilinguals' Arithmetic Solving

Performance differences that arise when bilinguals solve additions might also be due to the specific number word structure of the respective languages. To gauge the impact of the different two-digit number-naming systems used in French vs. German on arithmetic performance in the five different bilingualism proficiency groups, we investigated two aspects of the number words.

Firstly, we explored whether the particular *base-20 number word structure* used in French (but not in German) for numbers from 70 to 99 might impact arithmetic performances differentially across age-groups. Indeed, the number words under 70 follow the classical base-10 structure in both task languages, while the number words over 70 follow the base-10 structure in

German but not in French (where they follow the base-20 structure). We expected to find a general problem size effect in both languages because arithmetic problems with larger numbers are assumed to be more difficult to solve than arithmetic problems with smaller numbers (Groen and Parkman, 1972). But more interestingly, we also assumed that additions involving numbers over 70 would be specifically difficult in French because of the base-20 structure (Seron and Fayol, 1994). This specific difficulty should thus be especially pronounced at lower French proficiency levels.

Secondly, we aimed to understand whether and how the *order of tens and units in number words* (i.e., tens-units in French vs. units-tens in German) plays a role in bilinguals' addition performances. As Pixner et al. (2011) reported that the number naming system used in a two-digit number transcoding task modulated the type of errors, we analyzed which errors bilingual participants made on complex additions across the different presentation formats and languages. Given the contrasting positions of units and tens in German and French it is plausible that the same bilingual participant makes errors that predominantly pertain to distinct value positions depending on the language in which the calculation is performed.

Testing these predictions on our unique German-French bilingual sample will allow us to better understand the relation between language and arithmetic in bilinguals and how this relation evolves with increasing bilingual proficiency levels. To the best of our knowledge, there are currently no studies that systematically investigated how the influence of number word structure on arithmetical performance evolves as a function of language proficiency. Taken together these original data should also yield new insights into the role of language in arithmetic and number processing in general.

## Methods

### Participants

A total of 193 bilingual participants were recruited for the present study. The sample was composed of 36 pupils from grade 7 (21 females; mean age of 12.2 years;  $SD = 0.36$  years), 33 pupils from grade 8 (13 females; mean age of 13.2 years;  $SD = 0.58$  years), 35 pupils from grade 10 (15 females; mean age of 15.5 years;  $SD = 0.66$  years), 41 pupils from grade 11 (19 females; mean age of 16.4 years;  $SD = 0.72$  years) and 48 young adults (34 females; mean age of 22.4 years;  $SD = 2.67$  years).

All participants thus spoke Luxembourgish (an official language of Luxembourg which developed from a dialectal variant of German) or German as native language and attended the Luxembourgish school system in the highest academic track, which prepares for attending college and university. Moreover, all study participants (including the adults) had attended Luxembourgish primary school that starts with German as teaching language. From second grade of primary school on, all participants learned French as a second language. Importantly, students in grades 7 and 8 were taught mathematics in French, whereas students in grades 10 and 11 were not only taught mathematics but also all of their other courses in French (except the German and English language courses). Over the school years, relative

exposure and proficiency in French thus progressively increased and tended toward bilingualism with high proficiency levels in both German and French in the highest grades. Consequently, the adult group was composed of young adult participants who had become highly proficient German-French bilinguals through their education.

Native language(s), the number of years spent in Luxembourgish schools and linguistic background (under the form of self-assessment of language proficiency) were checked in a short questionnaire before starting the experiment in order to ensure that all participants also had similar exposures to languages in these respects. Adults received 20€ for their participation. Informed consent was obtained from all participants.

### Stimuli

Eighty-four two-operand addition problems were presented during the entire experiment. The set was composed of 28 one-digit simple additions (e.g.,  $4 + 2$ ) and 56 two-digit complex additions (e.g.,  $56 + 32$ ). This stimulus set was split in four blocks of additions to be allocated to both presentation formats of the problems and to both language sessions: visual and auditory presentation of the numbers in the German session and visual and auditory presentation of the numbers in the French session.

Simple additions were composed of two one-digit operands ranging from 2 to 9. We excluded  $+1$  additions and additions between the same operands (e.g.,  $7 + 7$ ), resulting in a range of solutions from 5 to 17. The simple additions with carry (additions with a solution of 10 or more) and without carry (additions with a solution below 10) were equally distributed across the four blocks of additions.

Complex additions were composed of two two-digit operands ranging from 12 to 86 in order to keep solutions below 100. We excluded all additions including a zero or ties. Furthermore, problems with a repetition of the same digit between the operands or between one of the operands and the solution were excluded, resulting in a range of solutions from 35 to 98. The requirement of a carry to be solved (with or without carry), the position of the larger operand (left vs. right in visual presentation; first vs. second in auditory presentation) and the problem size (small when the solution ranged between 30 and 69 or large when the solution ranged between 70 and 98) were taken into account in the repartition of complex additions in the four blocks. Indeed, each block contained seven problems with carry and seven problems without carry, and seven problems of small size and seven problems of large size. In other words, among the small problems, half of them contained a carry and half of them did not, and the same for the large problems, so that problems with and without carry were distributed equally among problems of different sizes within each block. The assignment of the blocks to a presentation format and a language was balanced through participants. For instance, block 1 was assigned to visual presentation of the French session for the first eight participants but the same block 1 was assigned to visual presentation of the German session for the next eight participants.

### Procedure

We ran the experiment on an Apple 13' Macbook using Pyscope X B57 (Cohen et al., 1993) where voice onset times of responses

**TABLE 1 | Mean duration of presentation of auditory additions in ms with standard deviation for each complexity level of the additions as a function of language.**

	Simple additions		Complex additions	
	<i>M</i>	( <i>SD</i> )	<i>M</i>	( <i>SD</i> )
German	1736	(80.1)	2582	(134.6)
French	1745	(169.8)	2574	(235.3)
Total	1740	(131.6)	2576	(191.8)

Mean presentation time differences between languages were not significant neither for simple additions,  $t_{(54)} = 0.239$ ;  $p = 0.812$ , nor for complex additions,  $t_{(108)} = 0.148$ ;  $p = 0.883$ .

were recorded with a voice key on the Iolab USB Button Box. As the voice key only recorded the response onset, the experimenter wrote the solutions down and pressed a key to start the next trial, which started after an inter-trial interval of 500 ms. The onset of the response time (RT)—measurement started when the stimulus presentation was completed.

In the visual presentation format, additions appeared on a white screen in black (Arial, font size 90) until participants responded. In the auditory presentation format, participants had to listen to the additions via headphones (in both ears). The length of auditory presentation was controlled between languages separately for simple and complex additions, so that the mean duration of auditory presentation did not differ between languages (see **Table 1**). In both presentation formats, participants had to respond orally by giving the solution in the microphone in the language of the task. This means that for auditory presentation of the additions, RT-measurement started at the offset of the second operand.

The testing was organized in two language sessions: participants performed both presentation formats first in one task language and then in the other. Order of presentation formats and task languages were counterbalanced between participants. Instructions and interaction with the experimenter remained in German or in French, according to the session. Participants were tested individually and were instructed to respond as accurately and as fast as possible. Seven training items preceded the 21 additions of each block. The entire experiment lasted about 50 min.

## Data Processing

### Effects of Calculation Complexity (Simple vs. Complex Additions)

In order to track the development of arithmetic problem solving in bilingual children and adults, correct response times (RTs) and correct response rates (CRs) during experimental tasks were collected at five different stages of language proficiency. Training items were not included in the dataset, and we also excluded RTs of all trials below or above three standard deviations from the mean of each participant and from the group mean. We excluded 4% of the trials in this way before analyzing the RTs.

We ran a preliminary analysis of variance (ANOVA) on the RTs and the CRs including all additions participants had to solve with Complexity<sub>2</sub> × Format<sub>2</sub> × Task language<sub>2</sub> as within-subject

factors and Age-group<sub>5</sub> as between-subject factor. The two levels of complexity were the simple one-digit operand vs. the complex two-digit operand addition problems; format referred to visual or auditory presentation of the additions; and task language was German or French (for instructions, presentation of the additions in the auditory format, and production of the answer). The age-group factor had the following levels: seventh graders, eighth graders, tenth graders, eleventh graders or young adults. The aim of this preliminary ANOVA was to see whether it was relevant to analyze both complexity levels (simple vs. complex addition) separately. Therefore, we only report results from the effect and interactions with the complexity factor. Then, we ran analyses of variance (ANOVA) on the RTs and the CRs separately for each type of additions: i.e., the simple one-digit additions and the complex two-digit additions. Within each ANOVA we used Format<sub>2</sub> × Task language<sub>2</sub> as within-subject factors and Age-group<sub>5</sub> as between-subject factor.

### Effects of Number Word Structure

To investigate how arithmetic performance is influenced by the different structures of number words in German and French we conducted two additional analyses. Firstly we tracked the impact of the particular base-20 number word structure used in French but not in German for numbers from 70 to 99 on the arithmetic performances across age-groups. Therefore, we introduced one more factor in the ANOVA on complex additions: the problem size. We categorized the items in two levels of problem size according to whether problems involved or not a number over 70. Indeed the number words under 70 follow the classical base-10 structure in both task languages, whereas the number words over 70 follow the base-10 structure in German but the base-20 structure French. We thus ran an ANOVA with Problem size<sub>2</sub> × Format<sub>2</sub> × Task language<sub>2</sub> as within-subject factors and with Age-group<sub>5</sub> as between-subject factor.

Secondly, we focused on the impact of the order of tens and units (i.e., ten-unit in French vs. unit-ten in German) in two-digit number words on arithmetic performances. We analyzed the type of errors participants made across different presentation formats and languages when solving complex additions involving on two-digit numbers. Within each task language and format, we listed the rate of errors (%) for which only the ten-digit was false (“ten-error,” i.e., 34 instead of 24) and inversely, the rate of errors for which only the unit-digit was false (“unit-error,” i.e., 34 instead of 35). Other types of errors were not included in the analyses because we found less than 2% of each type. We ran an ANOVA on these error rates with Error type<sub>2</sub> × Format<sub>2</sub> × Task language<sub>2</sub> as within-subject factors and with Age-group<sub>5</sub> as between-subject factor. The two levels of the error type factor corresponded to “ten-error” and “unit-error” and the level of the other factors were the same as in the previous analyses.

## Results

### Effects of Calculation Complexity (Simple vs. Complex Additions)

Preliminary ANOVA showed a strong effect of complexity on both RTs [ $F_{(1, 184)} = 893.961$ ;  $p < 0.001$ ;  $\eta^2 = 0.829$ ] and

CRs [ $F_{(1, 185)} = 510.891; p < 0.001; \eta^2 = 0.734$ ]. Both in RTs and CRs, complexity modulated effects of language [RTs:  $F_{(1, 184)} = 177.873; p < 0.001; \eta^2 = 0.492$ , CRs:  $F_{(1, 185)} = 64.294; p < 0.001; \eta^2 = 0.258$ ], format [RTs:  $F_{(1, 184)} = 43.034; p < 0.001; \eta^2 = 0.190$ , CRs:  $F_{(1, 185)} = 235.634; p < 0.001; \eta^2 = 0.560$ ] and age-group [RTs:  $F_{(4, 184)} = 12.740; p < 0.001; \eta^2 = 0.217$ , CRs:  $F_{(1, 185)} = 3.880; p = 0.005; \eta^2 = 0.077$ ]. We also observed a triple interaction between complexity, language and format [RTs:  $F_{(1, 184)} = 13.119; p < 0.001; \eta^2 = 0.067$ , CRs:  $F_{(1, 185)} = 9.768; p = 0.002; \eta^2 = 0.050$ ]. Only in RTs, there was also a significant triple interaction between complexity, language and age-group [RTs:  $F_{(4, 184)} = 4.610; p = 0.001; \eta^2 = 0.091$ ]. Since all factors of the preliminary ANOVA interacted with complexity, we will directly report below separate analyses and results for both complexity levels.

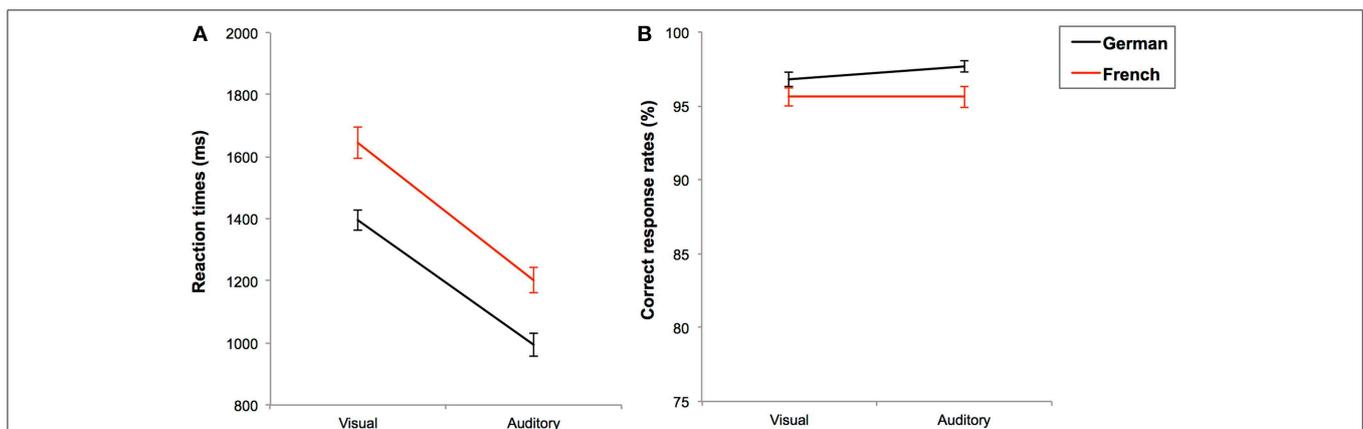
### Simple Additions

For the simple additions, overall mean RT was 1309 ms ( $SE = 32$  ms) and overall mean CR was 96.4% ( $SE = 0.3\%$ ). We found an age-group effect on RTs [ $F_{(4, 184)} = 12.710; p < 0.001; \eta^2 = 0.216$ ] and on CRs [ $F_{(4, 185)} = 3.038; p = 0.019; \eta^2 = 0.062$ ], as participants solved the simple additions faster and more accurately with increasing age-group (see **Table 2**). Furthermore, simple additions were performed faster when they were presented in auditory than in visual format,  $F_{(1, 184)} = 171.992; p < 0.001; \eta^2 = 0.483$ , but no difference between formats was observed in terms of CRs,  $F_{(1, 185)} = 0.737; p = 0.392; \eta^2 = 0.004$  (see **Figures 1A,B**). Thus, simple auditory additions were solved faster than visually presented ones, but correct response rates were similar for both formats.

Moreover, simple additions were performed faster in German than in French [RT:  $F_{(1, 184)} = 77.199; p < 0.001; \eta^2 = 0.296$ ],

**TABLE 2 | Means of reaction times (RT) in ms and correct response rates (CR) in % with standard errors for each complexity level of the additions (simple vs. complex) and the general mean performances as a function of age-group.**

Group	Simple additions		Complex additions		Total	
	<i>M</i>	( <i>SE</i> )	<i>M</i>	( <i>SE</i> )	<i>M</i>	( <i>SE</i> )
<b>RT</b>						
Seventh graders	1638	(74.5)	5774	(285.9)	3706	(173.7)
Eighth graders	1408	(77.9)	4533	(298.9)	2970	(181.6)
Tenth graders	1358	(75.6)	4377	(290.0)	2867	(176.2)
Eleventh graders	1145	(69.7)	3691	(264.1)	2383	(162.5)
Adults	994	(63.6)	3098	(244.1)	2046	(148.3)
Total	1309	(32.4)	4294	(124.0)	2794	(75.5)
<b>CR</b>						
Seventh graders	94.8	(0.7)	74.5	(1.9)	84.7	(1.1)
Eighth graders	95.7	(0.8)	77.0	(1.9)	86.4	(1.2)
Tenth graders	96.9	(0.7)	77.7	(1.9)	87.3	(1.1)
Eleventh graders	96.5	(0.7)	82.1	(1.7)	89.3	(0.1)
Adults	98.0	(0.6)	84.9	(1.6)	91.5	(0.9)
Total	96.4	(.03)	79.2	(0.8)	87.8	(0.5)



**FIGURE 1 | Mean reaction times in ms (A) and mean correct response rates in percentages (B) with standard errors for the simple additions in each task language (black line for German and red line for French) as a function of presentation format.**

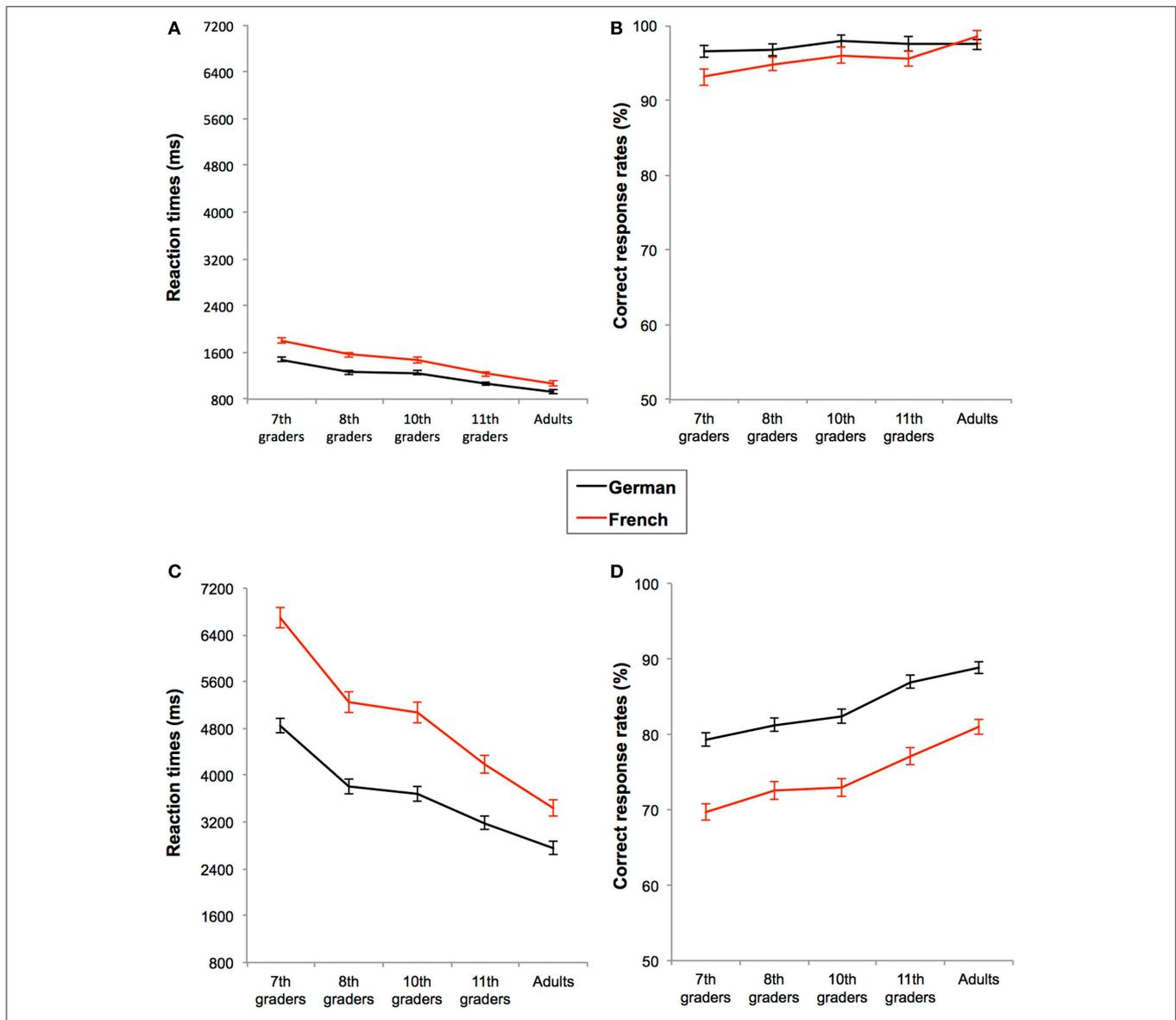
see **Figure 2A**. Participants also made fewer errors in German than in French [ $F_{(1, 185)} = 9.782$ ;  $p = 0.002$ ;  $\eta^2 = 0.050$ ], but this language effect on CRs was marginally modulated by the age-group (language  $\times$  age-group:  $F_{(4, 185)} = 2.234$ ;  $p = 0.067$ ;  $\eta^2 = 0.046$ ), see **Figure 2B**. We decomposed this interaction by separately running a  $\text{Format}_2 \times \text{Task language}_2$  ANOVA on CRs in each age-group. It appeared that only the seventh graders were less accurate in French than in German,  $F_{(1, 34)} = 4.074$ ;  $p = 0.050$ ;  $\eta^2 = 0.092$ , while all other age-groups performed with equal accuracy in both languages (all  $F$ 's  $< 1$  and  $p$ 's  $> 0.05$ ), see **Figure 2B**. No other interaction reached significance (all  $F$ 's  $< 1$  and  $p$ 's  $> 0.05$ ). In sum, participants solved simple additions faster in German than in French, but in both languages they performed the task faster when additions were presented in auditory

than in visual format. In terms of accuracy, additions presented in both languages and presentation formats were performed equally well, except that seventh graders were less accurate in French than in German.

### Complex Additions

For the complex additions, overall mean RT was 4294 ms ( $SE = 124$  ms) and overall mean CR was 79.2% ( $SE = 0.8\%$ ). We found an age-group effect on RTs [ $F_{(4, 185)} = 14.008$ ;  $p < 0.001$ ;  $\eta^2 = 0.232$ ] and on CRs [ $F_{(4, 185)} = 5.976$ ;  $p < 0.001$ ;  $\eta^2 = 0.114$ ], as participants from the older age-groups solved the complex additions faster and more accurately (see **Table 2**).

Regardless of task language complex additions were performed faster [ $F_{(1, 185)} = 4.997$ ;  $p = 0.027$ ;  $\eta^2 = 0.026$ ], and



**FIGURE 2 | Mean performances for simple additions (A, B) and for complex additions (C, D) in each task language (black line for German and red line for French) as a function of age-group. Performances are measured in reaction times (A, C) and in correct response rates (B, D). Bars represent standard errors.**

**TABLE 3 | Means of reaction times (RT) in ms and correct response rates (CR) in % with standard errors for each presentation format of the simple and complex additions as a function of age-group.**

Group	Simple additions				Complex additions			
	Visual		Auditory		Visual		Auditory	
	<i>M</i>	( <i>SE</i> )	<i>M</i>	( <i>SE</i> )	<i>M</i>	( <i>SE</i> )	<i>M</i>	( <i>SE</i> )
<b>RT</b>								
Seventh graders	1863	(86.9)	1414	(79.4)	5430	(260.2)	6118	(349.1)
Eighth graders	1577	(90.9)	1239	(83.0)	4485	(272.1)	4580	(365.1)
Tenth graders	1575	(88.2)	1142	(80.6)	4283	(263.9)	4470	(354.2)
Eleventh graders	1357	(81.3)	933	(74.3)	3682	(240.4)	3699	(322.6)
Adults	1231	(74.2)	757	(67.8)	3038	(222.1)	3158	(298.1)
Total	1520	(37.8)	1097	(34.5)	4183	(112.8)	4405	(151.5)
<b>CR</b>								
Seventh graders	94.6	(0.9)	95.1	(1.0)	86.2	(1.6)	62.8	(2.6)
Eighth graders	95.5	(0.9)	96.0	(1.0)	84.5	(1.6)	69.4	(2.7)
Tenth graders	96.6	(0.9)	97.3	(1.0)	85.2	(1.6)	70.2	(2.6)
Eleventh graders	96.3	(0.8)	96.7	(0.9)	87.3	(1.4)	76.8	(2.4)
Adults	97.9	(0.8)	98.1	(0.8)	89.1	(1.3)	80.8	(2.2)
Total	96.2	(0.4)	96.6	(0.4)	86.5	(0.7)	72.0	(1.1)

more accurately [ $F_{(1, 185)} = 245.736; p < 0.001; \eta^2 = 0.571$ ], when presented visually than in auditory format, see **Table 3**. Moreover, for CRs the format effect was modulated by age-group [format  $\times$  age-group interaction:  $F_{(4, 185)} = 8.365; p < 0.001; \eta^2 = 0.153$ ]. Decomposition of this interaction showed that participants became more accurate with age for auditory presented additions. However, CRs remained similar across age-group for visually presented additions. This led to a progressively smaller error rate difference between visual and auditory formats with age-group (see **Table 3**).

In general, complex additions were also performed faster and more accurately when the task language was German than when it was French, RT:  $F_{(1, 185)} = 201.922; p < 0.001; \eta^2 = 0.522$  and CR:  $F_{(1, 185)} = 113.630; p < 0.001; \eta^2 = 0.381$ , see **Figures 2C,D**. However, this language effect was modulated by the presentation format, both in terms of RTs,  $F_{(1, 185)} = 10.729; p = 0.001; \eta^2 = 0.055$ , and CRs,  $F_{(1, 185)} = 19.657; p < 0.001; \eta^2 = 0.096$ . Firstly, results from the pairwise comparisons on the RTs showed that even if complex additions were always performed faster in German than in French, the effect of the format (i.e., visual vs. auditory) was only significant for French,  $F_{(1, 185)} = 9.867; p = 0.002; \eta^2 = 0.051$ , but not for German,  $F_{(1, 185)} = 0.105; p = 0.746; \eta^2 = 0.001$ . Hence, auditory-presented additions were performed slower than visually presented additions only in French, see **Figure 3A**. Secondly, results from the pairwise comparisons on the CRs showed that language-related accuracy differences were larger in auditory,  $F_{(1, 185)} = 109.919; p < 0.001; \eta^2 = 0.373$ , than visual presentation format,  $F_{(1, 185)} = 28.536; p < 0.001; \eta^2 = 0.134$ , see **Figure 3B**.

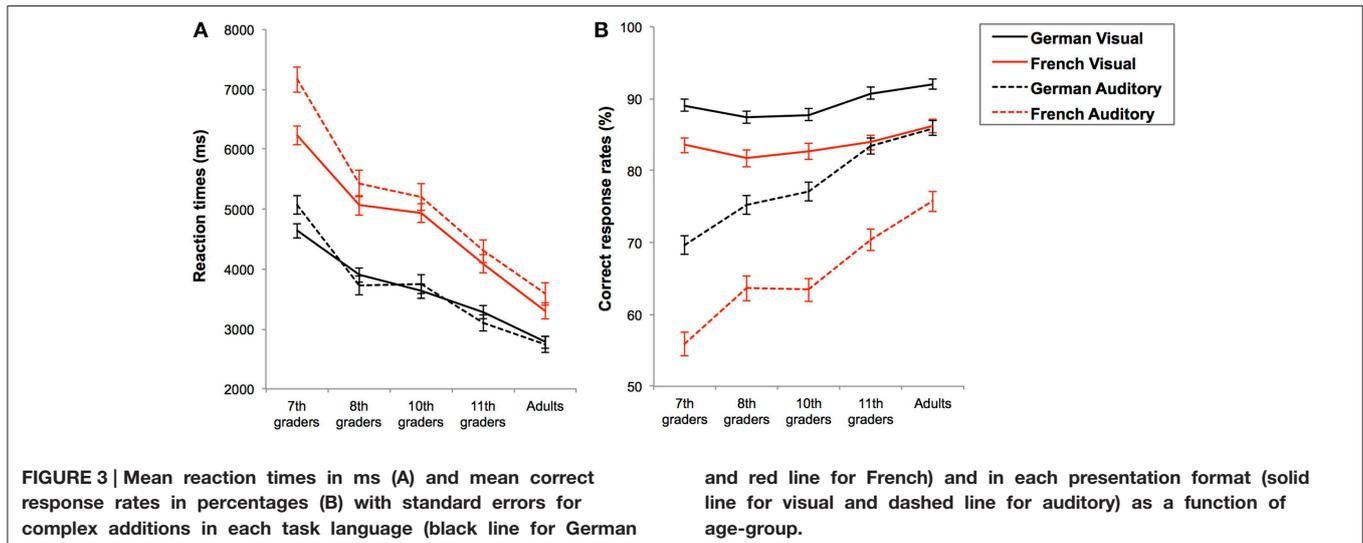
Thus, when task language was French, participants were slower for additions presented in auditory compared to visual format, but the presentation format did not modulate RTs in German. Additionally, additions of the German session were

always solved more accurately than additions of the French session and this effect of task-language was more pronounced for additions presented in auditory format. Finally, regardless of presentation format, task language also interacted with the age-group on the RTs,  $F_{(4, 187)} = 5.317; p < 0.001; \eta^2 = 0.103$ , but not on the CRs,  $F_{(4, 187)} = 0.194; p = 0.941; \eta^2 = 0.004$ . Indeed, response times of both language sessions became increasingly similar with age, see **Figure 2C**.

When considering the above analyses it appears that the variability was different across age groups. Levene’s test for homogeneity of variances across groups was indeed significant, as the younger groups performances were more heterogeneous than the older groups’ (see standard errors reported in **Table 2**). This characteristic of the data is typical for transversal developmental comparisons, but it might have impacted the above-mentioned results and masked some interactions between age groups and task-language and/or presentation format effects. To cancel any potential influences of variance heterogeneity we therefore re-conducted the same analyses after a standardization of the data per age-group. The results of this additional analysis are detailed in the Annex 1 of Supplementary Material.

To sum up results on both raw and standardized data, bilingual participants of all five age groups solved *simple additions* faster in German than in French. Moreover auditory format simple additions were performed faster than visual format additions in both languages<sup>1</sup>. In contrast age group impacted the accuracy of simple addition solving, as seventh graders were overall less accurate in French than in German. This finding was confirmed

<sup>1</sup>Direct differences of RTs between formats of presentation can only be interpreted as modality-related measurement differences because RT recording started at the end of stimulus presentation in auditory format and at the beginning of stimulus presentation in visual format.



by the z-score analyses, which revealed that only participants of the tenth grade onwards solved simple additions with equal accuracy in both languages, even if they remained always slightly faster in German than in French.

Furthermore, *complex additions* were performed faster and better in German than in French. Critically, age group impacted RT differences observed when bilingual participants solved complex additions in German compared to French. Nevertheless, additional results on z-scores showed that task-language effect on RTs did no longer interact with age-group after standardization of the data. Thus, the effect of task language on RTs and CRs remained constant proportionally across age-groups.

Concerning presentation format, even though the differences of RTs cannot be interpreted *per se*, results on CRs showed that participants made more errors in auditory format than in visual format, especially in French compared to German where the CRs difference between formats was smaller. And this effect interacted with age-group as participants became more accurate for auditory-presented additions with increasing age. This last interaction between format and age-group remained significant after standardization of the data, suggesting that participants' ability to solve auditory presented additions genuinely improves with age.

### Effects of Number Word Structure Base-10 vs. Base-20 Tens

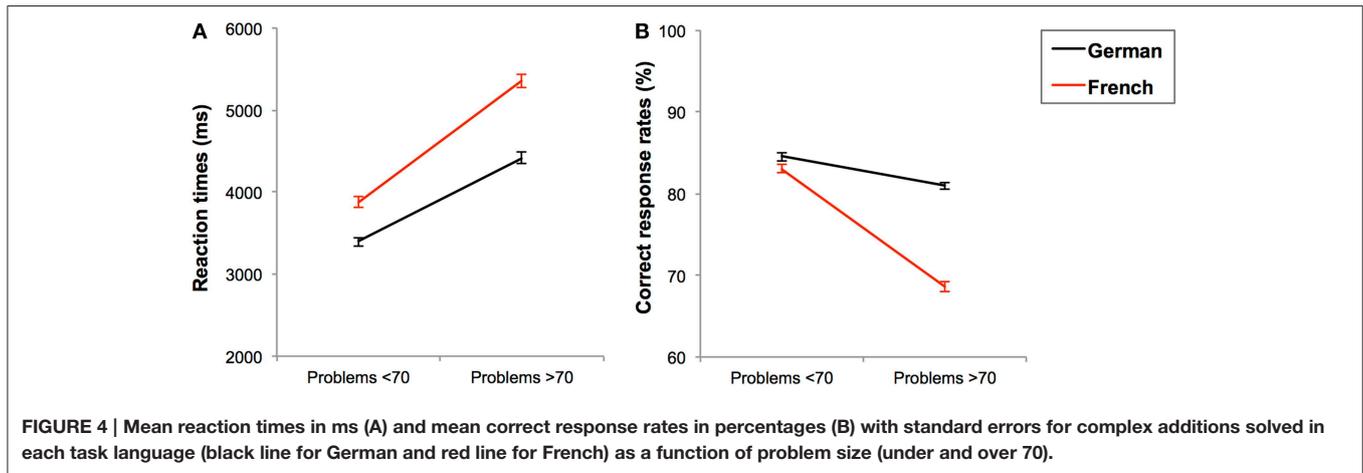
Here we only report effects and interactions involving the problem size factor because other effects and interactions were already explained in detail in section Effects of Calculation Complexity (Simple vs. Complex Additions). In general, we observed lower CRs [ $F_{(1, 185)} = 85.196$ ;  $p < 0.001$ ;  $\eta^2 = 0.315$ ] and slower RTs [ $F_{(1, 173)} = 151.138$ ;  $p < 0.001$ ;  $\eta^2 = 0.466$ ] with problems over 70 than with problems under 70. Moreover, problem size interacted with the task-language both in RTs,  $F_{(1, 173)} = 16.327$ ;  $p < 0.001$ ;  $\eta^2 = 0.086$ , and CRs,  $F_{(1, 185)} = 52.912$ ;  $p < 0.001$ ;  $\eta^2 = 0.222$ .

To decompose this interaction, we ran pairwise comparisons. The problem size effect on the RTs was larger when the task language was French [ $F_{(1, 173)} = 91.565$ ;  $p < 0.001$ ;  $\eta^2 = 0.346$ ] than when it was German [ $F_{(1, 173)} = 64.836$ ;  $p < 0.001$ ;  $\eta^2 = 0.273$ ], see **Figure 4A**. In terms of CRs, problem size effect was only significant in French [ $F_{(1, 185)} = 105.613$ ;  $p < 0.001$ ;  $\eta^2 = 0.363$ ] but not in German [ $F_{(1, 185)} = 2.878$ ;  $p = 0.091$ ;  $\eta^2 = 0.015$ ], see **Figure 4B**. Further, the difference in CRs between German and French was smaller in problems under 70 [ $F_{(1, 185)} = 11.763$ ;  $p = 0.001$ ;  $\eta^2 = 0.060$ ] than in problems over 70 [ $F_{(1, 185)} = 140.200$ ;  $p < 0.001$ ;  $\eta^2 = 0.431$ ]. Finally, the problem size factor did not interact with any other factor, not even the age-group, all  $F_s < 1$  and  $p_s > 0.1$ . Thus, task language strongly modulated the effect of problem size in the direction that problem size effects were more pronounced when the task was performed in French than in German.

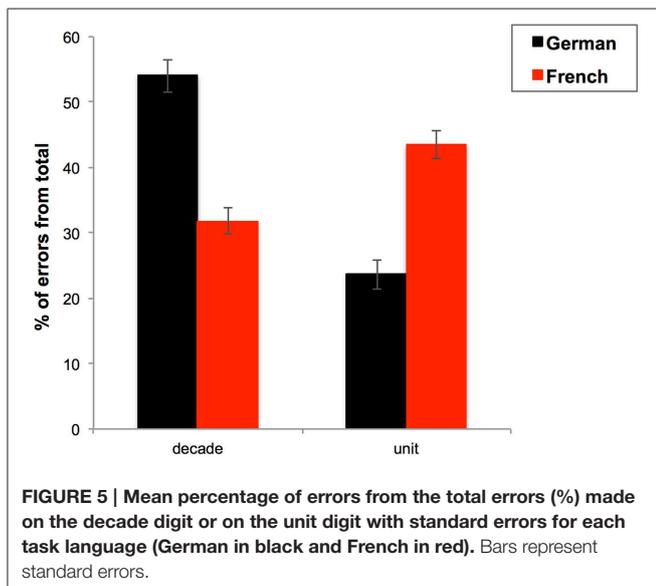
When considering the above analyses it appears that participants generally responded slower in French, which was also their less mastered language. Thus, the greater problem size effect found in French could also be due to participants' weaker French proficiency, independently of the structure of number words in this language. To rule out this alternative explanation, we re-conducted this analysis after a standardization of the data per language, see results in Annex 2 of Supplementary Material. In summary, interactions of language and problem size remained significant after standardization of the data per language, suggesting that differences of problem size effect observed between languages in raw data are not a consequence of bilinguals' differences between languages in terms of language mastery.

### Units-Tens vs. Tens-Units

In general, more errors were made on tens than units,  $F_{(1, 110)} = 10.283$ ;  $p = 0.002$ ;  $\eta^2 = 0.085$ . Moreover, the task language  $\times$  error type interaction was significant,  $F_{(1, 110)} = 56.194$ ;  $p < 0.001$ ;  $\eta^2 = 0.338$ , and pairwise comparisons showed that there were more errors on the tens than on the units when additions were presented in German,  $F_{(1, 110)} = 50.108$ ;  $p < 0.001$ ;



**FIGURE 4 |** Mean reaction times in ms (A) and mean correct response rates in percentages (B) with standard errors for complex additions solved in each task language (black line for German and red line for French) as a function of problem size (under and over 70).



**FIGURE 5 |** Mean percentage of errors from the total errors (%) made on the decade digit or on the unit digit with standard errors for each task language (German in black and French in red). Bars represent standard errors.

$\eta^2 = 0.313$ , but inversely, when additions were presented in French, there were more errors on the units than on the tens,  $F_{(1, 110)} = 9.594$ ;  $p = 0.002$ ;  $\eta^2 = 0.080$ . Additionally, there were more errors on the tens in German than in French,  $F_{(1, 110)} = 48.293$ ;  $p < 0.001$ ;  $\eta^2 = 0.305$ , and more errors on the units in French than in German,  $F_{(1, 110)} = 46.711$ ;  $p < 0.001$ ;  $\eta^2 = 0.298$ , see **Figure 5**.

Furthermore, the format of presentation interacted with task language,  $F_{(1, 110)} = 7.783$ ;  $p = 0.006$ ;  $\eta^2 = 0.066$ . Pairwise comparisons of German vs. French addition errors showed the same pattern of results in both presentation formation but more errors of both types were made in German than in French for the auditory presentation format,  $F_{(1, 110)} = 8.002$ ;  $p = 0.006$ ;  $\eta^2 = 0.068$ . In contrast, more errors of both types were made in visual than in auditory presentation format in French,  $F_{(1, 110)} = 12.791$ ;  $p = 0.001$ ;  $\eta^2 = 0.104$ . It should be noted that this last interaction does not change the conclusions yielded by the aforementioned complex additions results, as here the error rates

only referred to percentage of errors on unit vs. 10 digits in the incorrect solutions.

In summary, in German more errors were produced on the tens (e.g., “twenty” in “four-and-twenty”), whereas errors concerned predominantly the units in French (“four” in twenty-four”). This pattern of results was present for both presentation formats of the additions but was even more prominent in auditory format when task was performed in German and in visual format when task was performed in French.

## Discussion

To provide new insights into the question of bilingual’s arithmetic problem solving we tracked arithmetic performances in German-French bilinguals at five different stages of their bilingual development from adolescence to adulthood. Four age-groups of pupils attending secondary school and one group of young adults had to provide oral answers to simple (i.e., addends <10) and complex (i.e., addends >10) addition problems presented once in a visual format (Arabic digits) and once in an auditory format (spoken number words). Moreover, all participants performed experimental tasks both in German and French in two distinct language sessions. Task language had a direct influence on solving complex addition problems, whereas only much weaker language effects were observed when participants retrieved answers for simple additions. From adolescence to adulthood complex additions performance considerably improved in both German and French, with especially noteworthy gains of accuracy in auditory-presented calculations. Yet, for complex additions a substantial language-related advantage for German additions remained in highly proficient adult bilinguals both in accuracy and response times. In contrast, participants tended to retrieve simple additions comparably well in German and French with increasing bilingual proficiency. In addition, the specific number word structures of German and French also significantly impacted bilinguals’ arithmetic performance. Due to the base-20 structure of large French two-digit words, calculations with large numbers over 70 were less well-succeeded in French than German. Furthermore, the tendency to make errors

involving the second position of the number word led bilingual participants to produce more errors on the units when calculating in French and more decade-related errors in German. Firstly, we will discuss how language globally affected task performance and then separately consider simple and complex addition solving. Secondly, we will debate upon the effect of number word structure on bilinguals' arithmetic skills.

### Effects of Calculation Complexity

Overall additions were performed faster and with fewer errors in German than in French. This task language effect seemed to persist even in highly proficient adult bilinguals. As German was learned first by all participants it can be considered as their predominant language. In addition, it was also their language of arithmetic acquisition. Our results are consequently in agreement with the fact that (a) relative language predominance is known to promote arithmetic performance in bilinguals (Marsh and Maki, 1976; McClain and Huang, 1982; Frenck-Mestre and Vaid, 1993; Geary et al., 1993) and (b) bilingual adults solve numerical problems more proficiently in the language in which arithmetic was learned (Bernardo, 2001; Salillas and Wicha, 2012). The results also fit with the idea of non-selective language activation in bilinguals. Thus, lower arithmetic performances in French might also—at least partially—be due to less efficient access for French (in general) than for the predominant German (Bialystok, 2009; Kroll et al., 2013, 2014).

Nevertheless, a more nuanced picture emerged when considering separately how performance in simple and complex additions varied between the increasing language proficiency levels. With simple additions (e.g.,  $4 + 3 = 7$ ) seventh and eighth graders were still marginally less accurate in French than German. But all other participants from grade 10 and upward did not show any accuracy difference between German and French when solving simple additions. If arithmetic acquisition language alone would explain language-related differences in bilinguals' arithmetic performance, then we would have expected an advantage for simple additions in the German session persisting in all age groups. However, here we observed that after 3 years of math-classes in French (i.e., grade 10 and upwards) participants solved simple additions with equal accuracy levels in German and French. This suggests that in addition to the importance the language for arithmetic acquisition, the current language proficiency level modulated the ability to retrieve simple arithmetic facts. Once a certain proficiency level was reached in both of the bilinguals' languages, the initial advantage for solving simple additions in the language in which they had been acquired (i.e., German) no longer applied for accuracy rates, as participants attained ceiling performances for these very simple arithmetic problems. However, even in adults some response time differences between languages remained, though reduced in comparison to other age-groups.

In complex addition (e.g.,  $54 + 13 = 67$ ), language-related performance differences were more prominent as responses remained slower and less accurate in French than in German in all groups, which is consistent with the idea that complex additions require more processing steps and are therefore more likely to be influenced by language (Beishuizen, 1993). At first sight,

raw data analysis indicated that complex addition response times of both language sessions became increasingly similar with age. However, when group differences in variance were eliminated by data standardization it appeared that language-related performance differences in favor of German remained of similar importance across all age groups. Concerning German, even with mathematics taught in French during the entire secondary school years, we observed neither decrease nor stagnation of arithmetic performances in comparison to French. Thus, complex calculation proficiency in the first language (i.e., German) seems to pursue a continuous development independently of the language in which formal math education is taught.

Complex additions were also affected differentially by presentation format of the additions, whereas no substantial difference was observed in simple additions. Participants made always more mistakes with auditory-presented additions. But French still enhanced this effect, with participants making on average 34% ( $\pm 0.01\%$  SE) errors when computing auditory-presented complex additions (vs. 22% ( $\pm 0.01$  SE) errors in German). Over and above this interaction with task language, auditory-presented complex additions were succeeded less well than visually-presented ones. However, the auditory disadvantage gradually reduced with increasing age (even in standardization data). This relative improvement for auditory-presented additions that was specific to complex problems might be due to developmental trends in cognitive and verbal abilities combined with a prolonged exposure to complex addition solving and an increasing math expertise. Indeed, as attested by the ceiling performances observed in simple additions, all participants were perfectly skilled to retrieve arithmetic facts, coherently with the common observation that children usually achieve arithmetic fact retrieval around the age of 8 years onwards (Barrouillet and Fayol, 1998; Butterworth, 2005). Although, participants' performances on complex additions did not reach any ceiling and continued to improve across age-groups in both languages. These observations fit well with the idea that solutions for complex additions cannot be retrieved directly from memory, even in adults (Ashcraft, 1995).

For this type of complex arithmetic computation, factors such as procedural knowledge, planning and working memory are known to play critical roles (Fürst and Hitch, 2000). In auditory presentation format, the additional need to keep the heard addends in working memory may interfere with using the phonological loop in the computation process. Consequently, participants made more errors for auditory presented additions than visually presented additions. This format effect in complex additions was especially pronounced when performing the additions in French, i.e., a language that was relatively less proficient (LeFevre et al., 2001) and/or distinct from the language of arithmetic acquisition. These findings nicely highlight the involvement of language in the numerical processing underlying complex additions. If participants had simply computed the results in their first language (i.e., German) and then translated them to the output language (i.e., French) this would have affected performance similarly in both the visual and the auditory presentation formats. But contrary to this prediction performance specifically dropped when participants computed

auditory-presented complex additions in French. This may be due to the fact that arithmetic was learned in German or to globally weaker proficiency level in French. Due to the specific differences between number word structure in French and German languages, the interaction might also (at least partially) result from differences between French and German number naming systems. In the following paragraphs, we will further discuss the latter effects and their relation to arithmetic in German-French bilinguals.

### Effects of Number Word Structure

Languages differ in the way they construct two-digit number words (Campbell and Xue, 2001). This may directly influence bilinguals' addition skills and/or interact with other factors such as bilingual proficiency level and arithmetic acquisition language. Evaluating how arithmetic problem solving is influenced by number word structure in German-French bilinguals is particularly interesting because those both languages encounter two major differences in their number naming systems. Firstly, two-digit number words follow a unit-ten order in German (e.g., "24" = four-and-twenty) but a ten-unit order in French (e.g., "24" = twenty-four). Secondly, the 10 words for the numbers over 70 follow a base-10 structure in German (e.g., "72" = two-and-seventy) but a base-20 structure in French (e.g., "72" = sixty-twelve).

To characterize the effect of number word differences between languages on arithmetic performances, we conducted additional analyses on complex additions. Firstly, we focused on the base-10 vs. base-20 structure of large two-digit number words. Additions involving numbers under and over 70 were analyzed separately, since number words under 70 follow a base-10 structure in both language but number words over 70 follow a base-ten structure in German and a base-20 structure in French (e.g., "72" is pronounced as "sixty-twelve"). Not surprisingly, additions over 70 were solved overall slower than additions under 70 in both languages, confirming the classical problem size effect (Groen and Parkman, 1972). Nevertheless, the response time difference between additions under and over 70 was larger in French than in German. Moreover, in terms of accuracy, participants made more errors for additions over 70 than additions under 70 in French, but not in German where errors rates in additions under and over 70 were similar. Interestingly, these results were observed regardless of additions' presentation formats and participant groups. The latter observation demonstrates that the base-10 vs. base-20 effect is not modulated by bilingual proficiency groups. Nevertheless, it remains to be empirically determined whether specific difficulties for number words also occur in French-German bilinguals with French as first language. These findings confirm the early reports by Ellis and Hennelly (1980) that bilinguals' arithmetic skills are inevitably marked and modulated by the number word structure of the language in which they are currently calculating. In line with the present results, recent behavioral and electrophysiological studies indicate that these language-related characteristics might even impregnate basic number representations (Pixner et al., 2011; Salillas and Carreiras, 2014).

Secondly, we analyzed the type of errors participants made, namely whether more errors were made on the tens or on the

units across different languages, presentation formats, and bilingual proficiency groups. As noted above, the number word structure in French and German differs in terms of which digit is pronounced at first in two-digit number words (ten vs. unit). It appeared that participants systematically produced more errors on the ten digit (e.g., "2" in "24") when calculating in German and more errors on the unit digit (e.g., "4" in "24") in French. Again, the presentation format and the group of participants did not modify this result. Thus, independent of the calculation language, errors seem to predominantly concern digits holding the second position of the solution number. These findings elegantly show how a language-independent focus on the first segment of number words can lead to qualitatively distinct numerical outcomes within different language contexts. Taken at face value, they imply that making calculation errors while computing prices in the range between 18 and 100 will become more expensive for a German- than for a French-speaking person.

### General Considerations

Literature provides divergent conclusions about the level at which bilinguals' different languages are involved in number processing and about the language in which bilinguals actually solve arithmetic problems. Many factors such as age of acquisition of the second language, language of teaching during school years and currently used language seem to determine the use of the language during arithmetic problem solving in bilinguals (Bernardo, 2001; Campbell and Epp, 2004; Salillas and Wicha, 2012). Investigating arithmetic performance in Luxembourgish adolescents and young adults who become highly proficient German-French bilinguals through the school system offered the rare opportunity to study large groups of bilingual participants at different bilingual proficiency levels who are homogeneously composed with respect to the previous factors.

Our findings obtained with German-French bilinguals at five distinct levels of bilingual proficiency extend the current knowledge by confirming that language plays a critical role in the computations underlying complex addition (i.e., operands above "10") at all bilingual stages. Participants' skills in computing additions in both German and French improved steadily with increasing bilingual proficiency levels from grade 7 to young adulthood. Nevertheless, participants of all age groups solved complex German additions faster and more accurately than French ones. This German advantage remained although mathematics is taught in French during the entire secondary school years. It is probably due to the fact that German is participants' first school language and their arithmetic acquisition language and that complex additions are not automatized enough to be free of any language help along the solving process. In contrast, simple addition facts (i.e., operands below "10") were accessed more directly and similarly in both languages, especially at later stages of second language acquisition. Indeed accuracy levels for simple additions were similar in French and German from grade 10 upwards, while their response times got closer. Thus, highly proficient bilinguals tend to be able to retrieve addition facts similarly in both languages suggesting that bilinguals' arithmetic fact retrieval may become either independent from the verbal code or

automatized enough in different languages' verbal codes to lead to similar performances (Campbell and Epp, 2004).

The second part of our study explored the role of number word structure in bilinguals' arithmetic performance. German-French bilinguals indeed speak two languages that are characterized by inverted ten-unit structures of two-digit number words (unit-ten vs. ten-unit number words) and with different constructions of tens over 70 (base-10 vs. base-20). Consequently the full effect of number word structure on arithmetic computation could be highlighted optimally in this type of bilingual population. When additions were computed in French, specific response-delays and error-increases were observed for calculations involving number words over 70. Moreover, results from error analyses showed that participants of all age groups always committed more errors related to the digit that occurred in second position in the number word, i.e., tens in German and units in French. Taken together, both differences in German vs. French number word structures (two-digit words with base 10 vs. 20 and direct vs. inverted digit order) seemed to play a role in arithmetic processing at all bilingual proficiency stages.

In conclusion, the present study demonstrates that both (a) language proficiency levels and (b) number word structure affect addition solving performances in bilinguals. This leads to

the conclusion that arithmetic significantly relies on language processes, especially in complex computations. Further studies will be needed to generalize the present findings to other number processing tasks (e.g., magnitude comparison), other arithmetic operations (e.g., subtraction, multiplication,) and other tasks with number words (e.g., math word problems).

## Acknowledgments

This work was supported by the Langnum CORE project funded by the Luxembourgish Fund for Scientific Research (FNR, Luxembourg). The authors declare APA ethical standards were followed in the conduct of this study. Authors gratefully thank all the participants for their collaboration to the study, the school principals who gave the permission to recruit pupil participants and the students who helped to collect the data.

## Supplementary Material

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fpsyg.2015.00265/abstract>

## References

- Altarriba, J., and Basnight-Brown, D. M. (2007). Methodological considerations in performing semantic- and translation-priming experiments across languages. *Behav. Res. Methods* 39, 1–18. doi: 10.3758/BF03192839
- Ashcraft, M. H. (1992). Cognitive arithmetic: a review of data and theory. *Cognition* 44, 75–106.
- Ashcraft, M. H. (1995). Cognitive psychology and simple arithmetic: a review and summary of new directions. *Math. Cogn.* 1, 3–34.
- Baddeley, A. (1992). Working memory. *Science* 255, 556–559.
- Baldo, J. V., and Dronkers, N. F. (2007). Neural correlates of arithmetic and language comprehension: a common substrate? *Neuropsychologia* 45, 229–235. doi: 10.1016/j.neuropsychologia.2006.07.014
- Barrouillet, P., and Fayol, M. (1998). From algorithmic computing to direct retrieval: evidence from number and alphabetic arithmetic in children and adults. *Mem. Cognit.* 26, 355–368.
- Basso, A., Burgio, F., and Caporali, A. (2000). Acalculia, aphasia and spatial disorders in left and right brain-damaged patients. *Cortex* 36, 265–280. doi: 10.1016/S0010-9452(08)70528-8
- Basso, A., Caporali, A., and Faglioni, P. (2005). Spontaneous recovery from acalculia. *J. Int. Neuropsychol. Soc.* 11, 99–107. doi: 10.1017/S1355617705050113
- Beishuizen, M. (1993). Mental strategies and materials or models for addition and subtraction up to 100 in Dutch second grades. *J. Res. Math. Educ.* 24, 294.
- Benn, Y., Zheng, Y., Wilkinson, I. D., Siegal, M., and Varley, R. (2012). Language in calculation: a core mechanism? *Neuropsychologia* 50, 1–10. doi: 10.1016/j.neuropsychologia.2011.09.045
- Bernardo, A. B. (2001). Asymmetric activation of number codes in bilinguals: further evidence for the encoding complex model of number processing. *Mem. Cognit.* 29, 968–976. doi: 10.3758/BF03195759
- Bialystok, E. (2009). Bilingualism: the good, the bad, and the indifferent. *Biling. Lang. Cogn.* 12, 3. doi: 10.1017/S1366728908003477
- Brybaert, M., Fias, W., and Noël, M.-P. (1998). The Whorfian hypothesis and numerical cognition: is 'twenty-four' processed in the same way as 'four-and-twenty'? *Cognition* 66, 51–77.
- Butterworth, B. (2005). The development of arithmetical abilities. *J. Child Psychol. Psychiatry* 46, 3–18. doi: 10.1111/j.1469-7610.2004.00374.x
- Campbell, J. I. (1994). Architectures for numerical cognition. *Cognition* 53, 1–44. doi: 10.1016/0010-0277(94)90075-2
- Campbell, J. I., and Epp, L. J. (2004). An encoding-complex approach to numerical cognition in Chinese-English bilinguals. *Can. J. Exp. Psychol.* 58, 229–244. doi: 10.1037/h0087447
- Campbell, J. I., and Xue, Q. (2001). Cognitive arithmetic across cultures. *J. Exp. Psychol. Gen.* 130, 299–315. doi: 10.1037/0096-3445.130.2.299
- Cappelletti, M., Butterworth, B., and Kopelman, M. (2001). Sparing numerical abilities in a case of semantic dementia. *Neuropsychologia* 39, 1224–1239. doi: 10.1016/S0028-3932(01)00035-5
- Chen, Z.-Y., Cowell, P. E., Varley, R., and Wang, Y.-C. (2009). A cross-language study of verbal and visuospatial working memory span. *J. Clin. Exp. Neuropsychol.* 31, 385–391. doi: 10.1080/13803390802195195
- Cohen, J., MacWhinney, B., Flatt, M., and Provost, J. (1993). PsyScope: an interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behav. Res. Methods Instrum. Comput.* 25, 257–271.
- Cohen, L., Dehaene, S., Chochon, F., Lehéricy, S., and Naccache, L. (2000). Language and calculation within the parietal lobe: a combined cognitive, anatomical and fMRI study. *Neuropsychologia* 38, 1426–1440. doi: 10.1016/S0028-3932(00)00038-5
- Colomé, À., Laka, I., and Sebastián-Gallés, N. (2010). Language effects in addition: how you say it counts. *Q. J. Exp. Psychol.* 63, 965–983. doi: 10.1080/17470210903134377
- Dehaene, S., and Cohen, L. (1995). Towards an anatomical and functional model of number processing. *Math. Cogn.* 1, 83–120.
- Dehaene, S., Spelke, E., Pinel, P., Stanescu-Cosson, R., and Tsivkin, S. (1999). Sources of mathematical thinking: behavioral and brain-imaging evidence. *Science* 284, 970–974.
- Delazer, M., Domahs, F., Bartha, L., Brenneis, C., Lochy, A., Trieb, T., et al. (2003). Learning complex arithmetic—an fMRI study. *Cogn. Brain Res.* 18, 76–88. doi: 10.1016/j.cogbrainres.2003.09.005
- Delazer, M., Girelli, L., Semenza, C., and Denes, G. (1999). Numerical skills and aphasia. *J. Int. Neuropsychol. Soc.* 5, 213–221.
- Ellis, N. C., and Hennessey, R. A. (1980). A bilingual word-length effect: implications for intelligence testing and the relative ease of mental calculation in Welsh and English. *Br. J. Psychol.* 71, 43–51.

- Fayol, M., and Thevenot, C. (2012). The use of procedural knowledge in simple addition and subtraction problems. *Cognition* 123, 392–403. doi: 10.1016/j.cognition.2012.02.008
- Frenc-Mestre, C., and Vaid, J. (1993). Activation of number facts in bilinguals. *Mem. Cognit.* 21, 809–818.
- Fürst, A. J., and Hitch, G. J. (2000). Separate roles for executive and phonological components of working memory in mental arithmetic. *Mem. Cognit.* 28, 774–782. doi: 10.3758/BF03198412
- Fuson, K. C., Richards, J., and Briars, D. J. (1982). “The acquisition and elaboration of the number word sequence,” in *Children’s Logical and Mathematical Cognition*, ed C. J. Brainerd (New York, NY: Springer), 33–92.
- Geary, D. C., Cormier, P., Goggin, J. P., Estrada, P., and Lunn, M. C. E. (1993). Mental arithmetic: a componential analysis of speed-of-processing across monolingual, weak bilingual, and strong bilingual adults. *Int. J. Psychol.* 28, 185–201.
- Göbel, S. M., Moeller, K., Pixner, S., Kaufmann, L., and Nuerk, H.-C. (2014). Language affects symbolic arithmetic in children: the case of number word inversion. *J. Exp. Child Psychol.* 119, 17–25. doi: 10.1016/j.jecp.2013.10.001
- Gordon, P. (2004). Numerical cognition without words: evidence from Amazonia. *Science* 306, 496–499. doi: 10.1126/science.1094492
- Groen, G. J., and Parkman, J. M. (1972). A chronometric analysis of simple addition. *Psychol. Rev.* 79, 329–343.
- Gruber, O., Indefrey, P., Steinmetz, H., and Kleinschmidt, A. (2001). Dissociating neural correlates of cognitive components in mental calculation. *Cereb. Cortex* 11, 350–359. doi: 10.1093/cercor/11.4.350
- Hitch, G. J. (1978). The role of short-term working memory in mental arithmetic. *Cogn. Psychol.* 10, 302–323.
- Klessinger, N., Szczerbinski, M., and Varley, R. (2012). The role of number words: the phonological length effect in multidigit addition. *Mem. Cognit.* 40, 1289–1302. doi: 10.3758/s13421-012-0228-y
- Kroll, J. F., Bobb, S. C., and Hoshino, N. (2014). Two languages in mind: bilingualism as a tool to investigate language, cognition, and the brain. *Curr. Dir. Psychol. Sci.* 23, 159–163. doi: 10.1177/0963721414528511
- Kroll, J. F., Gullifer, J. W., and Rossi, E. (2013). The multilingual lexicon: the cognitive and neural basis of lexical comprehension and production in two or more languages. *Annu. Rev. Appl. Linguist.* 33, 102–127. doi: 10.1017/S0267190513000111
- Le Corre, M., and Carey, S. (2007). One, two, three, four, nothing more: an investigation of the conceptual sources of the verbal counting principles. *Cognition* 105, 395–438. doi: 10.1016/j.cognition.2006.10.005
- LeFevre, J.-A., Lei, Q., Smith-Chant, B. L., and Mullins, D. B. (2001). Multiplication by eye and by ear for Chinese-speaking and English-speaking adults. *Can. J. Exp. Psychol.* 55, 277–284. doi: 10.1037/h0087374
- Lemer, C., Dehaene, S., Spelke, E., and Cohen, L. (2003). Approximate quantities and exact number words: dissociable systems. *Neuropsychologia* 41, 1942–1958. doi: 10.1016/S0028-3932(03)00123-4
- Lin, J.-F. L., Imada, T., and Kuhl, P. K. (2011). Mental addition in bilinguals: an fMRI study of task-related and performance-related activation. *Cereb. Cortex* 22, 1851–1861. doi: 10.1093/cercor/bhr263
- Logie, R. H., Gilhooly, K. J., and Wynn, V. (1994). Counting on working memory in arithmetic problem solving. *Mem. Cognit.* 22, 395–410.
- Macizo, P., Herrera, A., Román, P., and Martín, M. C. (2010). Second language acquisition influences the processing of number words. *Proc. Soc. Behav. Sci.* 9, 1128–1134. doi: 10.1016/j.sbspro.2010.12.295
- Marsh, L. G., and Maki, R. H. (1976). Efficiency of arithmetic operations in bilinguals as a function of language. *Mem. Cognit.* 4, 459–464. doi: 10.3758/BF03213203
- McClain, L., and Huang, J. Y. S. (1982). Speed of simple arithmetic in bilinguals. *Mem. Cognit.* 10, 591–596. doi: 10.3758/BF03202441
- McCloskey, M. (1992). Cognitive mechanisms in numerical processing: evidence from acquired dyscalculia. *Cognition* 44, 107–157. doi: 10.1016/0010-0277(92)90055-2
- McCloskey, M., Caramazza, A., and Basili, A. (1985). Cognitive mechanisms in number processing and calculation: evidence from dyscalculia. *Brain Cogn.* 4, 171–196. doi: 10.1016/0278-2626(85)90069-7
- Noel, M.-P., and Seron, X. (1993). Arabic number reading deficit: a single case study or when 236 is read (2306) and judged superior to 1258. *Cogn. Neuropsychol.* 10, 317–339. doi: 10.1080/02643299308253467
- Nuerk, H., Weger, U., and Willmes, K. (2005). Language effects in magnitude comparison: small, but not irrelevant. *Brain Lang.* 92, 262–277. doi: 10.1016/j.bandl.2004.06.107
- Pesenti, M., Thioux, M., Seron, X., and Volder, A. D. (2000). Neuroanatomical Substrates of Arabic Number Processing, Numerical Comparison, and Simple Addition: A PET Study. *J. Cogn. Neurosci.* 12, 461–479. doi: 10.1162/089892900562273
- Pica, P., Lemer, C., Izard, V., and Dehaene, S. (2004). Exact and approximate arithmetic in an Amazonian Indigene group. *Science* 306, 499–503. doi: 10.1126/science.1102085
- Pixner, S., Zuber, J., Heömanová, V., Kaufmann, L., Nuerk, H.-C., and Moeller, K. (2011). One language, two number-word systems and many problems: numerical cognition in the Czech language. *Res. Dev. Disabil.* 32, 2683–2689. doi: 10.1016/j.ridd.2011.06.004
- Rossor, M. N., Warrington, E. K., and Cipolotti, L. (1995). The isolation of a calculation skills. *J. Neurol.* 242, 78–81. doi: 10.1007/BF00887820
- Saalbach, H., Eckstein, D., Andri, N., Hobi, R., and Grabner, R. H. (2013). When language of instruction and language of application differ: cognitive costs of bilingual mathematics learning. *Learn. Instr.* 26, 36–44. doi: 10.1016/j.learninstruc.2013.01.002
- Salillas, E., and Carreiras, M. (2014). Core number representations are shaped by language. *Cortex* 52, 1–11. doi: 10.1016/j.cortex.2013.12.009
- Salillas, E., and Wicha, N. Y. (2012). Early learning shapes the memory networks for arithmetic: evidence from brain potentials in bilinguals. *Psychol. Sci.* 23, 745–755. doi: 10.1177/0956797612446347
- Seron, X., and Fayol, M. (1994). Number transcoding in children: a functional analysis. *Br. J. Dev. Psychol.* 12, 281–300. doi: 10.1111/j.2044-835X.1994.tb00635.x
- Spelke, E., and Tsivkin, S. (2001). Language and number: a bilingual training study. *Cognition* 78, 45–88. doi: 10.1016/S0010-0277(00)00108-6
- Stanescu-Cosson, R., Pinel, P., van de Moortele, P., Le Bihan, D., Cohen, L., and Dehaene, S. (2000). Understanding dissociations in dyscalculia: a brain imaging study of the impact of number size on the cerebral networks for exact and approximate calculation. *Brain* 123, 2240–2255. doi: 10.1093/brain/123.11.2240
- Van Rinsveld, A., Ugen, S., Schiltz, C., Landerl, K., and Brunner, M. (2012). “How does language affect numerical processing? The role of the teaching language on two-digit magnitude judgments in German-French bilingual 7- to 11-graders,” *Poster Session Presented at EARLI SIG 22 Neuroscience and Education 2012 Meeting* (London).
- Venkatraman, V., Siong, S. C., Chee, M. W., and Ansari, D. (2006). Effect of language switching on arithmetic: a bilingual fMRI study. *J. Cogn. Neurosci.* 18, 64–74. doi: 10.1162/089892906775250030
- Wang, Y., Lin, L., Kuhl, P., and Hirsch, J. (2007). Mathematical and linguistic processing differs between native and second languages: an fMRI study. *Brain Imaging Behav.* 1, 68–82. doi: 10.1007/s11682-007-9007-y
- Xu, F., and Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition* 74, B1–B11. doi: 10.1016/S0010-0277(99)00066-9
- Zago, L., and Tzourio-Mazoyer, N. (2002). Distinguishing visuospatial working memory and complex mental calculation areas within the parietal lobes. *Neurosci. Lett.* 331, 45–49. doi: 10.1016/S0304-3940(02)00833-9

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Van Rinsveld, Brunner, Landerl, Schiltz and Ugen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Number word structure in first and second language influences arithmetic skills

Anat Prior\*, Michal Katz, Islam Mahajna and Orly Rubinsten

Edmond J. Safra Brain Research Center for the Study of Learning Disabilities, University of Haifa, Haifa, Israel

## OPEN ACCESS

### Edited by:

Yvette Renee Harris,  
Miami University, USA

### Reviewed by:

Silvia Pixner,  
UMIT-The Health and Life Science  
University, Austria  
Yukari Okamoto,  
University of California,  
Santa Barbara, USA

### \*Correspondence:

Anat Prior,  
Edmond J. Safra Brain Research  
Center for the Study of Learning  
Disabilities, Faculty of Education,  
University of Haifa, Mount Carmel,  
Haifa 31905, Israel  
aprior@edu.haifa.ac.il

### Specialty section:

This article was submitted to  
Developmental Psychology, a section  
of the journal *Frontiers in Psychology*

**Received:** 13 September 2014

**Accepted:** 23 February 2015

**Published:** 17 March 2015

### Citation:

Prior A, Katz M, Mahajna I and  
Rubinsten O (2015) Number word  
structure in first and second language  
influences arithmetic skills.  
*Front. Psychol.* 6:266.  
doi: 10.3389/fpsyg.2015.00266

Languages differ in how they represent numerical information, and specifically whether the verbal notation of numbers follows the same order as the symbolic notation (in non-inverted languages, e.g., Hebrew, “25, twenty-five”) or whether the two notations diverge (in inverted languages, e.g., Arabic, “25, five-and-twenty”). We examined how the structure of number-words affects how arithmetic operations are processed by bilingual speakers of an inverted and a non-inverted language. We examined Arabic-Hebrew bilinguals’ performance in the first language, L1 (inverted) and in the second language, L2 (non-inverted). Their performance was compared to that of Hebrew L1 speakers, who do not speak an inverted language. Participants judged the accuracy of addition problems presented aurally in L1, aurally in L2 or in visual symbolic notation. Problems were presented such that they matched or did not match the structure of number words in the language. Arabic-Hebrew bilinguals demonstrated both flexibility in processing and adaptation to the language of aural-verbal presentation – they were more accurate for the inverted order of presentation in Arabic, but more accurate for non-inverted order of presentation in Hebrew, thus exhibiting the same pattern found for native Hebrew speakers. In addition, whereas native Hebrew speakers preferred the non-inverted order in visual symbolic presentation as well, the Arabic-Hebrew bilinguals showed enhanced flexibility, without a significant preference for one order over the other, in either speed or accuracy. These findings suggest that arithmetic processing is sensitive to the linguistic representations of number words. Moreover, bilinguals exposed to inverted and non-inverted languages showed influence of both systems, and enhanced flexibility in processing. Thus, the L1 does not seem to have exclusive power in shaping numerical mental representations, but rather the system remains open to influences from a later learned L2.

**Keywords:** L1, L2, bilingualism, number processing, addition

## Introduction

Bilingual speakers have control of two languages and hence raise important questions regarding language and cognitive representations and processing. Such questions include the degree to which two languages are represented or processed independently versus interactively (e.g., Kroll and Stewart, 1994; Costa, 2005), as well as the impact of language on cognitive representations more generally. The fact that languages differ in their structural properties provides an

elegant method of investigating to what degree cognitive representations, in the current case in the numerical domain, are uniquely shaped by the native language (L1) or rather remain open to influences from a later acquired second language (L2). This question has been previously addressed in the domain of semantic/conceptual representations. Thus, Jiang (2002, 2004) proposed a model according to which the conceptual system is mostly shaped by the L1, except in highly proficient bilinguals. In contrast, Degani et al. (2011) demonstrated sensitivity of semantic processing to L2 lexical properties, even in unbalanced moderately proficient bilinguals (see also Cook, 2003; Laufer, 2003; Wolff and Ventura, 2009). The role of the L1 vs. the L2 in shaping representations and processing has also been investigated in the numerical domain (Gelman and Butterworth, 2005), again leading to conflicting results.

For example, Macizo et al. (2010) showed that the processing of number words in one language was not modulated by the way bilinguals processed number words in their alternative language, which differed in the structure of number words. In contrast, it has been consistently shown that even proficient L2 speakers resort to their L1 to perform mathematical operations (e.g., Spelke and Tsivkin, 2001). In addition, Salillas and Carreiras (2014) recently demonstrated specific influences of the structure of the language in which early math instruction occurred on the processing of numerical information in highly proficient balanced bilinguals.

These conflicting results raise the question to what degree people can learn to process numerical information according to the structure of their L2, when it differs markedly from that of the L1, specifically when they are highly proficient L2 speakers. The current work extends this controversial line of research by investigation the impact of both the L1 and the L2 on numerical representation and calculation in highly proficient bilinguals, whose languages differ in the structure of number words.

This question is of central importance, because arithmetic processing in monolinguals is closely linked to language (Dehaene, 1992; McCloskey, 1992; Campbell, 1994; De Smedt et al., 2010; Archibald et al., 2013). Thus, it has been shown that the four basic operations (addition, subtraction, multiplication, and division) are learned in school with different emphasis on quantity manipulations and on linguistic skills (Dehaene and Cohen, 1995; Delazer et al., 2006; Ischebeck et al., 2006), with multiplication and addition being retrieved from verbal memory but subtraction and division requiring manipulation of quantities (e.g., Zhou et al., 2006). In general, it has been suggested that with advanced age and practice, counting and using quantity knowledge to achieve an outcome is replaced as the strategy of choice by memory retrieval, similar to the way words are retrieved from the verbal lexicon, at least in cases of addition and multiplication (e.g., Delazer et al., 2006; Ischebeck et al., 2006). However, it should be noted that Pesenti et al. (2000) and Venkatraman et al. (2005) did not find any language-related frontal activations for symbolic exact arithmetic involving simple addition problems, suggesting that different strategies, other than retrieval from verbal memory, may be in use.

These findings lead to fascinating questions concerning the cognitive mechanisms underlying mathematical operations in

proficient bilinguals, especially when information is presented in the L2. For example, when doing arithmetic in the L2, do bilinguals rely on the linguistic structure of that language, and how do these processes interact with the L1? These questions address the fundamental issue of whether human cognitive capacities related to the L1 and the L2 employ a shared or independent cognitive system. Numerical knowledge acts as a natural and ecological laboratory for the study of L1/L2 interactions, as bilinguals have three sets of symbols to represent the same semantic concept: written or symbolic digits (3), L1 number words (e.g., *shalosh* in Hebrew or *talate* in Arabic) and L2 number words. This makes it possible not only to study translation from L1 to L2 and from L2 to L1 but also from a common semantic meaning to written or verbal forms in either language. In the current study we extend the examination of bilingual cross-language interaction by asking whether the structure of number words in one language is modulated by the way bilinguals process numbers in the alternative language.

Languages differ in the structure of number words and how they are used, and such differences can shape the way in which speakers of a certain language process numbers. Thus, several studies set out to examine the idea that variability in mathematics performance may be related to differences in the cognitive organization of numbers that is affected by number-word characteristics of a language. Thus, number words in Chinese, Japanese, and Korean are congruent with the traditional base 10 numeration system, such that the spoken number corresponds exactly to the implied quantity represented in the written form (i.e., the number 49 is written in character symbols as four-10s-nine). Number words in English, on the other hand, may lack the elements of 10s and ones that are contained in them (i.e., the number 12, twelve). Miura et al. (1988) found that whereas first grade native speakers of English preferred to use a collection of unit blocks to represent numbers, speakers of base-10 languages more frequently used a construction of 10s and ones, in correspondence with the linguistic structure (see also Fuson and Kwon, 1992; Miller et al., 1995; Geary et al., 1996).

In a more recent study, Colomé et al. (2010) compared Italian and Catalan speakers. Italian is a base-10 language while Catalan number-words are constructed by combining multiples of 20 with units or with teens (e.g., the verbal representation of 35 is “twenty and fifteen”). Their results showed a consistent difference between the two groups in their preference toward a certain number-word structure when solving problems verbally and when typing their answers in Arabic numerals. The researchers concluded that language differences in the structure of number-words play a role when solving addition problems.

The current study focuses on the property of inversion, coined by Zuber et al. (2009) to describe the situation when the order of the symbolic and verbal notation of a number are inverted. For example, the number “25” in inverted languages is pronounced as “five and twenty.” The inversion property affects all two-digit numbers from 21 to 98, repeats for the 10,000s, and is a feature of various languages such as Arabic, Danish, Dutch, and German. There is evidence showing that children who speak languages with inversion have difficulty in basic numerical transcoding tasks, namely the ability to translate numerals from one form

to another, such as the Arabic notation “27” to verbal notation “seven and twenty” (e.g., Pixner et al., 2011; Imbo et al., 2014). Difficulty in such tasks, probably due to the multiple inversions required in representing two-digit numbers, consequently leads to working memory overload (Zuber et al., 2009).

In adults, Brysbaert et al. (1998) tested the theory that numerical addition is based on language processes by comparing French and Dutch-speaking participants solving addition problems such as  $20 + 4$  and  $4 + 20$  (unit + decade) and  $21 + 5$  and  $5 + 21$  (unit + decade-unit) presented either as Arabic numerals or as number words. The French participants solved operations like “ $20 + 4$ ” and “ $21 + 5$ ” faster than their counterparts in the other order, both when presented with Arabic digits and with number words. The Dutch participants differed in their performance; operations like “ $20 + 4$ ” were performed faster in this order only when presented in the numeric format, but no differences between “ $20 + 4$ ” and “ $4 + 20$ ” were found for the verbal format. When the operation consisted of decade-unit + unit ( $21 + 5$ ), they were faster to answer in the inverse order ( $5 + 21$ ). The results demonstrated some differences in preference for order of operations based on language when the questions were presented in written verbal form and the participants were asked to respond verbally. On the other hand, the two groups did not differ significantly in their responses when asked to type the answers numerically. The authors concluded that the numerical system is largely autonomous of the language system.

However, a later study by Nuerk et al. (2005) compared English and German speakers’ performance in magnitude comparison. Two numbers were presented above each other on a computer screen and participants were asked to determine which number was larger. This study showed influence of the inversion property in a unit-decade compatibility effect. This compatibility effect is found when two-digit Arabic numbers are compared, such that cases where separate decade and unit comparisons lead to the same decision (e.g., 32\_47; in this case  $3 < 4$  and  $2 < 7$ ) are processed faster than incompatible trials (e.g., 37\_52; in this case  $3 < 5$ , but  $7 > 2$ ). According to McCloskey (1992), there may be separate mental number line representations for decades and units which, in turn, may be separately processed in two-digit number comparison. If this is true, comparing a pair of incompatible numbers could be a more difficult and lengthy process than comparing a pair of compatible numbers.

In the above mentioned study, Nuerk et al. (2005) investigated the generality of the compatibility effect by comparing English and German speakers. They found that while for native German speakers the compatibility effect is much larger for large unit distances than for small unit distances, for native English speakers the compatibility effect is larger for small decade distances than that of the German speakers. Moreover, large unit distances and small decade distances led to disproportionately more errors for English participants but not for German participants. The authors therefore concluded that decade distance seemed to determine responses in English speakers, while overall distance was the most important predictor for German speakers, particularly when dealing with written number words. Thus, the lexical representation in a language influences magnitude comparison even when numbers are presented in a non-linguistic format.

A recent study conducted by Macizo and Herrera (2010) strengthens this conclusion by testing Spanish speakers’ number processing when presented with two-digit number words in reverse form (unit-decade order, e.g., five-and-twenty). In each trial, one number word was presented above the other in the center of the screen and the participants had to select the larger of the two numbers. Based on the effects of the decade distance and the compatibility effect, the results showed that only decade distance was a significant predictor for difference in reaction time (RT). The authors concluded that speakers of non-inverted languages have learned a language-dependent process for analyzing written numbers in which decades have a major role regardless of the position in which they are presented experimentally. These findings reinforce the theory that the spoken language does in fact affect the way in which numbers are processed when presented in both numeric and verbal form.

To date, there is only a handful of studies that have taken a close look at number processing in bilinguals who speak both an inverted language and a non-inverted language. These studies have suggested that bilinguals process two-digit number words selectively in their L1 and L2 and that they do not seem to transcode number words from their L2 into Arabic number format. In other words, most studies have found that the processing of number words in the L1 does not influence the way bilinguals process number words in their L2 (Macizo et al., 2010, 2011). Macizo et al. (2010) examined the way Italian/German bilinguals performed a number comparison task by presenting them with compatible and incompatible number–word pairs in their two languages. Participants were faster when presented with compatible pairs than incompatible pairs in German, while they were slower when presented with compatible pairs than incompatible pairs in Italian. The authors concluded that bilingual speakers are not bound to the number-structure of their L1 and the relative reliance on the decade and unit values differ depending on the language of presentation; when processing number–words in an inverted language, they rely on the unit values, and when processing number–words in a non-inverted language, they rely more on the decade values.

A more recent study by Macizo et al. (2011), also investigated between-language influences by comparing Spanish/English and German/English bilinguals’ performance on a number comparison task. Their results show that both bilingual groups presented a reverse compatibility effect when performing the comparison task in the L2 (a non-inverted language) but differed in the way they processed L1 numbers. A reverse compatibility effect was observed in the L1 Spanish task for the Spanish/English bilinguals (an expected pattern for a non-inverted language), and a regular compatibility effect was observed in the L1 German task for the German/English bilinguals (an expected pattern for an inverted language). The finding that bilinguals processed two-digit number words selectively in their L1 and L2 means that bilinguals are influenced by the language of presentation and process numbers according to the expected pattern for each language.

Taking such recent findings into account, the question that remains unanswered is whether or not cross-language influences exist in other numerical processing tasks, namely in arithmetic calculation. We investigate this question by presenting

Arabic/Hebrew bilinguals and native Hebrew speakers with addition problems composed both in visual-symbolic notation and in aural-verbal presentation. Similar to Brysbaert et al. (1998) we manipulated the order in which the elements of the addition problems were presented ( $20 + 5$  vs.  $5 + 20$ ) such that they did or did not match the structure of number words in the language. To our knowledge, this was the first study to use number words in an aural-verbal format instead of presenting number words in a written format. Thus, the current study will test the influence of language on number processing by examining the effect of the structure of number words in a language on processing addition problems, as well as the susceptibility of speakers of inverted and non-inverted languages to decade and unit numerical values.

We are particularly interested in whether the organization of numerical processing is determined by one's L1, which in this case is also the language of math instruction, or whether it is open to influences from the L2 as well. If the former is true, the performance of Arabic speakers in both aural-verbal presentation and in visual-symbolic presentation should reflect the inversion property of their L1. However, if the latter is true there are two possible patterns: they might show different preferences depending on the language in which the problem is presented. The second option is that Arabic-Hebrew bilinguals in the current study might show enhanced flexibility in processing, such that they become less sensitive overall to differences between presentations that match inverted or non-inverted structures.

## Materials and Methods

### Participants

Sixty three students from the University of Haifa participated in the study: 31 Arabic-Hebrew bilinguals (22 women, mean age 22) and 32 Hebrew-English bilinguals (20 women, mean age 26). Participants were recruited through flyers and online ads. Participants gave informed consent and were paid 30 NIS an hour (45–60 NIS in total). The study was approved by the research ethics committee of the University of Haifa. All participants included in the study reported no history of language and/or numerical disabilities.

### Materials

#### Language Experience and Proficiency Questionnaire (LEAP-Q)

The LEAP-Q (Marian et al., 2007) is a computerized self-report questionnaire that gathers information regarding participants' language background and abilities in all the languages they speak. The questionnaire includes questions regarding age of acquisition of languages, oral and written self-rated proficiency in all the languages a participant speaks, and the percent of time each language is used. The questionnaire was written in Hebrew and all participants were encouraged to ask questions if a portion of the questionnaire was unclear to them.

#### Arithmetic Two-Minute Test

Participants' mathematical automaticity skills were assessed using the Arithmetic Two-Minute test (Openhin-Bitton and Breznitz,

unpublished). This task consists of 80 simple arithmetic calculation problems, including the four basic math operations (addition, subtraction, multiplication, and division). The problems were presented in four columns, 20 problems for each basic math operation. Participants were instructed to solve as many problems as possible, from all four types, in 2 min. Total time, accuracy and correct responses per minute were scored.

### Working Memory Test

Memory performance was assessed using a computerized N-Back task (Owen et al., 2005), comprised of digit and spatial memory subsets. In both tasks, a sequence of digits or square locations was displayed on the computer screen and participants indicated when the current stimulus was identical to the stimulus that appeared on the previous trial by pressing on the "space" bar. There were 60–75 steps in each task (totaling 135 steps), 15 of which included target stimuli. Each trial started with a fixation point for 250 ms, a black screen for 500 ms, a stimulus for 500 ms, and a black screen for one second. Digit span was assessed using six digits (1, 2, 3, 4, 5, 6), and spatial memory was assessed using six different square locations on the computer screen. Participants could respond once the stimulus appeared or after 1 s. In addition, 5 s breaks were provided every 24 trials.

### Experimental Task: Verifying Addition Problems

Participants responded to addition problems presented to them in three formats: visual-symbolic (Arabic numerals), aural-verbal in the L1, and aural-verbal in the L2 (see **Table 1**). In order to balance the design, Hebrew speaking participants also completed an aural-verbal block in English, their L2. However, because the structure of number words does not differ between Hebrew and English, this block was not theoretically relevant, and therefore results were not analyzed.

Problems were presented with answers, and participants indicated by button press if the equation was correct or not. RT and accuracy of responses were recorded. All critical problems were comprised of the addition of a round decade number and a single unit number (e.g.,  $20 + 5 = 25$ ). Addition problems were constructed using three numerical ranges (20–29, 40–49, and 70–79). Elements of the problem could be presented such that they matched or did not match the order of number words in participants' language. The order manipulation was implemented across both aural-verbal and visual-symbolic presentation. Across participants each problem appeared in both the Match and the Non-match condition.

#### Match

The structure of the verbal representation of the problem *matches the structure of number words in the language*; i.e., "five plus twenty equals five and twenty" or " $5 + 20 = 25$ " for Arabic and "twenty plus five equals twenty five" or " $20 + 5 = 25$ " for Hebrew.

#### Non-match

The structure of the verbal representation of the problem *does not match* the structure of number words in the language; i.e., "twenty plus five equals five and twenty" or " $20 + 5 = 25$ " for Arabic, and "five plus twenty equals twenty five" or " $5 + 20 = 25$ " for Hebrew.

**TABLE 1 | Demonstration of experimental materials, by problem type and condition.**

		Match	Non-match
Correct problems	<i>Arabic aural-verbal</i>	Five plus twenty equals five and twenty	Twenty plus five equals five and twenty
	<i>Arabic visual-symbolic</i>	$5 + 20 = 25$	$20 + 5 = 25$
	<i>Hebrew aural-verbal</i>	Twenty plus five equals twenty-five	Five plus twenty equals twenty-five
	<i>Hebrew visual-symbolic</i>	$20 + 5 = 25$	$5 + 20 = 25$
Incorrect decade	<i>Arabic aural-verbal</i>	Five plus twenty equals thirty-five	Twenty plus five equals thirty-five
	<i>Arabic visual-symbolic</i>	$5 + 20 = 35$	$20 + 5 = 35$
	<i>Hebrew aural-verbal</i>	Twenty plus five equals thirty-five	Five plus twenty equals thirty-five
	<i>Hebrew visual-symbolic</i>	$20 + 5 = 35$	$5 + 20 = 35$
Incorrect unit	<i>Arabic aural</i>	Five plus twenty equals twenty-four	Twenty plus five equals twenty-four
	<i>Arabic numeric</i>	$5 + 20 = 24$	$20 + 5 = 24$
	<i>Hebrew aural</i>	Twenty plus five equals twenty four	Five plus twenty equals twenty four
	<i>Hebrew numeric</i>	$20 + 5 = 24$	$5 + 20 = 24$

Although verbal representations in the table are written in English to illustrate the order of elements in the problems, in the actual experiment all verbal materials were recorded in Hebrew or Arabic. In addition, all visual-symbolic problems were presented in Arabic (not Indian) numerals.

For each addition problem *correct* and *incorrect* responses were constructed. Incorrect answers consisted of an error either in the *units* or in the *decades*:

#### **Incorrect unit**

The wrong answer was in the same decade of the correct answer, but the unit value was different. If the numeral unity was under 5, it was replaced by a number between 5 and 9 at random; if the unit number was above 5, it was replaced by a number between 0 and 4 at random (see **Table 1**).

#### **Incorrect decade**

The wrong answer shared the same unit of the correct answer, but the decade value was different. Each group of decades was divided into two sub-groups: units under 5 and units above 5. In each sub-group, the decades were changed with a smaller value (minus 1) or greater value (plus 1) at random (see **Table 1**).

Finally, two types of filler addition problems were added to the list. The first type included problems from the second decade (11–19) of similar structure to the critical items. The second type of filler items were problems which did not match the structure of number words in either of the languages; e.g., “twenty three plus four equals twenty seven.” These problems were included in the experiment in order to provide the participants with a list of diversified problems and so that they do not pick up on a pattern of the first two types of problems. The filler problems could also include carry procedures. However, since this type of problems is not relevant for the theoretical questions presented in this study, they were not further analyzed.

When all stimuli were constructed, three comparable lists each containing 96 items were created. Each list included 24 items in the Match condition (12 correct, 6 incorrect Decade, 6 incorrect Unit); 24 items in the Non-match condition (12 correct, 6 incorrect Decade, 6 incorrect Unit) and 48 filler items (24 correct and 24 incorrect). All three lists were orally recorded in Arabic, Hebrew, and English, by a native speaking female of each language, respectively. Each problem and each answer was saved in separate sound files, played consecutively to participants. This allowed randomization of presentation order across participants,

and also allowed us to measure response RT from the onset of the answer, leading to more accurate assessment of performance.

#### **Procedure**

The tasks were divided into two 1-hour sessions. The first session included the LEAP-Q, the Two-Minute Test, and the Working Memory task. The second session included the experimental task of verifying addition problems. All computerized tasks were programmed in E-Prime, and the participants sat approximately 60 cm from the screen.

#### **Experimental Task Presentation**

##### **Aural-verbal blocks**

Each block started with written instructions in the language of the following block. Participants were instructed to respond as quickly and as accurately as possible.

Addition problems were presented through headphones, and did not appear on the screen, though participants responded using a computer keyboard. Each trial started with a fixation cross for 400 ms, followed by a blank screen for 150 ms, after which the problem was presented aurally while a green dot appeared in the center of the screen. The green dot remained on the screen until the participants responded. Participants used their index finger to press the right key for a correct answer or the left key for an incorrect answer. After responding, a red circle appeared in the center of the screen and participants pressed a key to initiate the following trial, to ensure that all participants had the same allotted response time. Each language block included 96 trials, and participants were given two short breaks during the block.

The instructions were followed by a practice block including 18 addition problems (nine problems per language). Participants were given feedback on their performance in the practice block. The experimental block, however, did not provide the participants with feedback on their performance.

##### **Visual-symbolic block**

Addition problems including answers were presented at the center of the screen. Each trial started with a fixation cross for 400 ms,

then a blank screen for 150 ms, after which the addition problem was presented centrally in Arabic numerals until participants responded with the right key if the problem was correct and with the left key if it was incorrect. Responses were followed by a red circle appearing in the middle of the screen, and participants pressed a key to initiate the next trial. The experimental block was preceded by a practice block of nine addition problems, for which participants received feedback.

Arabic speaking participants completed one list aurally in Arabic, one list aurally in Hebrew, and one list visually. Hebrew speaking participants completed one list aurally in Hebrew, one list aurally in English, and one list visually. The assignment of list to presentation condition was counterbalanced across participants, as were the order of visual vs. aural presentation, and the order of L1/L2 within the aural presentation. Within each list, item presentation was randomized for each participant. The 96 items in each list were randomly divided into three blocks, each containing 32 items. Participants were given breaks between blocks.

## Results

### Background Variables

The group performance in the background variables is presented in **Table 2**. The performance of the Arabic and Hebrew speakers was compared in working memory (N-back task), language background (LEAP-Q) and arithmetic abilities (Two-Minute arithmetic task). The Arabic speakers were significantly younger than the Hebrew speakers,  $t(60) = 5.32, p < 0.001$ . However, there was no significant difference between the groups in years of education ( $p = 0.15$ ).

Additionally, there was no significant difference between the two groups when performing the arithmetic two-minute test, ( $p = 0.66$ ). In other words, the Arabic speaking participants and the Hebrew speaking participants did not differ significantly in the number of arithmetic problems solved correctly in a two-minute span.

The working memory task, which required the participants to recall numerical and spatial stimulus 1 or 2 steps back, revealed a main effect of participant group, because Arabic speaker had shorter RTs than Hebrew speakers across all conditions,  $F(1,61) = 4.44, p < 0.05$ . However, both groups were equally accurate, again across all conditions,  $F(1,61) = 1.24, p = 0.27$ . Previous research has shown that accuracy in working memory

tasks is a more sensitive index of individual differences in working memory (Unsworth and Engle, 2008). Therefore, we do not further analyze the speed differences between the participant groups.

### Experimental Tasks – Addition Problems

In order to address the theoretical issue of the impact of number word structure on numerical processing, we conducted three main comparisons. In the processing of aural-verbal problems we first compared the performance of Arabic speakers in Arabic (the L1, an inverted language) and Hebrew (the L2, a non-inverted language). Then, we compared the performance of Hebrew speaking and Arabic speaking participants in their performance on Hebrew aural-verbal problems. This comparison allowed us to investigate whether speakers of an inverted L1 might process a non-inverted language differently than native speakers of a non-inverted L1. Finally, we compare the performance of the two participant groups on their responses to visual-symbolic problems. An important aspect of the two comparisons across participant groups is that they were based on the exact same stimuli for all participants.

### Arabic Speakers, L1/L2 Aural Presentation

To compare the performance of native Arabic speakers in L1 and L2, we conducted a three-way repeated-measures ANOVA on accuracy rates, and on mean RTs for correct responses. Within participant variables were Presentation Language (Arabic, Hebrew), Order (Match, Non-match to the structure of number words in the language of presentation), and Correctness (correct, incorrect Unit, incorrect Decade).

In the analysis of RTs, there was a main effect of presentation language  $F(1,28) = 42.5, p < 0.001, \eta = 0.6$ , because participants were faster to respond to addition problems in Arabic, the L1, than in Hebrew, the L2 (**Table 3**). Although participants were numerically faster to respond to problems that matched the structure of number words in the relevant language (inverted in Arabic, non-inverted in Hebrew), this difference did not reach statistical significance,  $F(1,28) = 2.1, p = 0.16$ . This finding is noteworthy in that it demonstrates that Arabic speaking participants were not sensitive to order of presentation, and regardless of whether they were listening to problems in the L1 or the L2 they were equally able to respond to problems presented in inverted or non-inverted order (see **Table 3**). Finally, the two-way interaction between presentation language and correctness was significant,  $F(2,56) = 19.1, p < 0.01, \eta = 0.7$ . This interaction is driven by the fact that in Arabic, participants were faster to respond to problems with an incorrect unit, whereas in Hebrew

**TABLE 2 | Means (SD) of participant characteristics.**

	Native Arabic speakers <i>N</i> = 31	Native Hebrew speakers <i>N</i> = 32
Age*	21.65 (2.4)	25.73 (2.9)
L1 self-rated proficiency	9.71 (0.49)	9.85 (0.31)
L2 self-rated proficiency	7.71 (1.37)	7.47 (1.88)
L2 age of acquisition	8.90 (1.7)	7.26 (3.15)
Participant years of education	14.63 (1.96)	14.35 (1.7)

\*Means significantly different at  $p < 0.01$ .

**TABLE 3 | Mean RTs (SD) for aural-verbal addition problems, by language and by order.**

	Native Arabic speakers in Arabic (L1)	Native Arabic speakers in Hebrew (L2)	Native Hebrew speakers in Hebrew (L1)
Inverted	1297 (50)	1697 (60)	1655 (64)
Non-inverted	1313 (66)	1635 (58)	1560 (57)

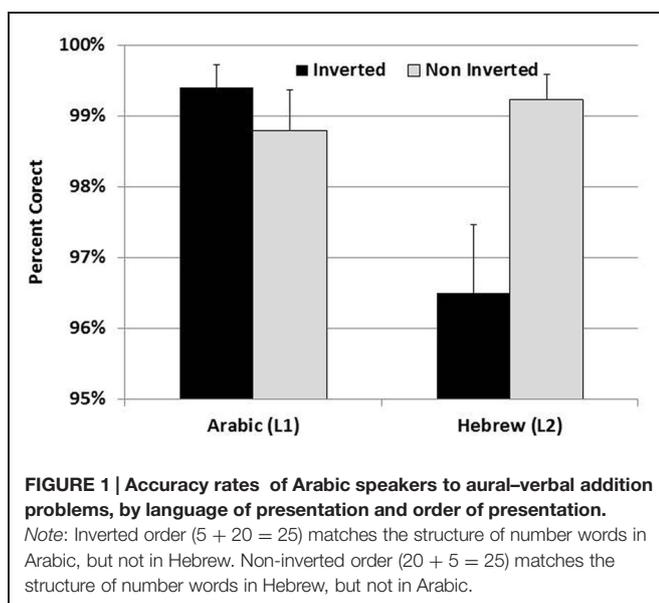
they were faster to respond to problems with an incorrect decade. Because in Arabic the unit information becomes available first in aural presentation (five-and-twenty) whereas in Hebrew the decade information becomes available first (twenty-and-five) this pattern is expected. No other main effects or interactions were significant.

In the accuracy analysis there was a significant main effect of presentation language,  $F(1,28) = 4.90, p < 0.05, \eta = 0.15$ , because participants were more accurate overall in the L1 than in the L2. In addition, there was a main effect of Order,  $F(1,28) = 7.7, p < 0.01, \eta = 0.2$ , because participants were more accurate to judge addition problems adhering to the structure of number words in the language of presentation, than to problems that did not match the structure of number words (see **Figure 1**). Importantly, the effect of Order was stable across both languages of presentation (namely, the interaction between Order and Language was not significant), indicating that in Arabic participants were more accurate in judging problems presented in the inverted order, whereas in Hebrew they were more accurate in judging problems presented in the non-inverted order. This shows flexibility and adaptability of processing preferences to the language of presentation.

### Comparing Hebrew and Arabic Speakers on Aural-Verbal Presentation in Hebrew

To compare the performance of native Hebrew and native Arabic speakers in responding to aural-verbal addition problems presented in Hebrew, we conducted a three-way mixed design ANOVA, on reaction times and accuracy (**Table 3**). Within-participant variables were Correctness (correct, incorrect-Unit, incorrect-Decade), Order (Match, Non-match to the structure of number words in the native language), and the between participant variable was native language group (Arabic, Hebrew).

Analysis of RTs to Hebrew aural presentation revealed a significant main effect of Order,  $F(1,58) = 5.9, p < 0.05, \eta = 0.1$ ,



because participants were faster to respond to addition problems that match the structure of number words in Hebrew, than to problems that do not match this structure. The two-way interaction between Order and Language Group was not significant,  $F < 1$ , showing that native Hebrew and native Arabic participants showed very similar patterns of performance and sensitivity to the order manipulation. This finding aligns with the pattern reported above, comparing the accuracy of performance of the native Arabic speakers in Arabic and in Hebrew.

Although native Hebrew speakers, performing the task in their L1, were numerically faster than native Arabic speakers performing the task in their L2 (mean RTs of 1607 and 1637 ms, respectively), this difference was not statistically significant,  $F < 1$ . The main effect of correctness was significant,  $F(1,58) = 7.1, p < 0.05, \eta = 0.35$ , because participants were slower to respond to problems with incorrect units ( $m = 1736$ ) than to correct problems ( $m = 1622$ ) or to problems with incorrect decades ( $m = 1551$ ). Again, we interpret this pattern as a result of the time at which information becomes available as the answer to the problem unfolds aurally. No other interactions were significant.

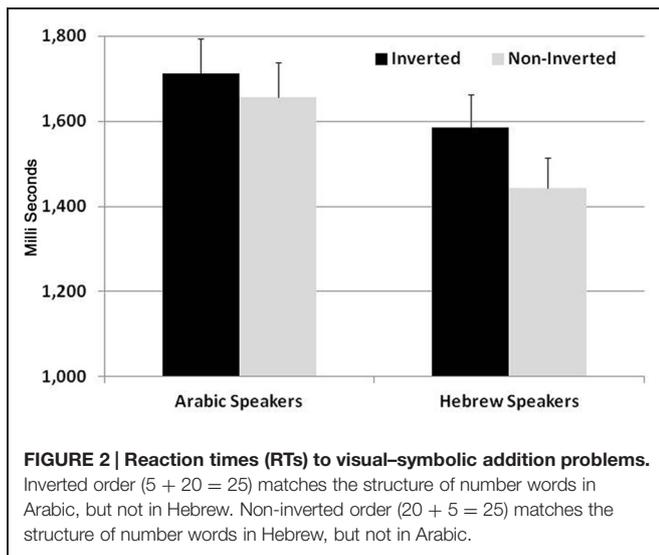
The analysis of accuracy rates again revealed a significant main effect of Order,  $F(1,58) = 6.6, p < 0.05, \eta = 0.1$ , because all participants were more accurate to judge addition problems that matched the structure of number words in Hebrew than problem that did not match this structure. Crucially, the effect of Order did not interact with Language Group, demonstrating that this preference was shared by both native Arabic and native Hebrew speakers. This is the same pattern that was reported above for the RTs. There were no other significant main effects or interactions.

### Comparing Hebrew and Arabic Speakers on Visual-Symbolic Presentation

To compare the performance of native Hebrew and native Arabic speakers in responding to visual-symbolic addition problems, we conducted a three-way mixed design ANOVA, on reaction times and accuracy (see **Figure 2**). Within-participant variables were Correctness (correct, incorrect-Unit, incorrect-Decade), Order (Match, Non-match to the structure of number words in the native language), and the between participant variable was native language group (Arabic, Hebrew).

In the analysis of RTs there was a significant main effect of Correctness,  $F(1,60) = 8.6, p < 0.01, \eta = 0.2$ . Participants were faster to respond to correct than to incorrect problems. There was also a significant two-way interaction between Order and Language group,  $F(1,60) = 6.7, p < 0.05, \eta = 0.1$ . Follow up comparisons showed that whereas native Hebrew speakers were significantly faster to respond to problems matching the structure of number words in Hebrew than to non-matching problems [ $t(30) = 2.7, p < 0.01$ ], native Arabic speakers did not show sensitivity to the order manipulation,  $t(30) < 1$ . No other main effects or interactions were significant.

In the analysis of accuracy rates, the only significant finding was a three-way interaction between Order, Correctness, and Language group,  $F(2,120) = 4.2, p < 0.05, \eta = 0.1$ . Follow up comparisons showed that for Arabic speakers there were no significant effects in accuracy for either Order of presentation or Correctness (all  $F < 1$ ). Conversely, for Hebrew speakers there



was a significant interaction between Order and Correctness,  $F(2,60) = 4.3$ ,  $p < 0.05$ , because they showed lower accuracy rates for problems with incorrect units presented in the non-matching order.

## Discussion

The present study examined whether adult university students show a preference for processing addition problems presented in an order that matches the structure of number words in their native language. Furthermore, we investigated the permeability of numerical processing to the structure of number words in the L2, especially when it differs markedly from the L1. We found that native Hebrew speakers, whose L2 (English) shares the same non-inverted structure of number words as the L1, have a marked preference both in aural-verbal presentation and in visual-symbolic presentation for addition problems presented in an order that matches the familiar structure of number words. Conversely, we found that Arabic-Hebrew bilinguals showed more flexibility in their patterns of performance, though the patterns revealed by the data were somewhat more complex. Thus, when comparing the performance of Arabic-Hebrew bilinguals across their two languages and for visual-symbolic problems, they did not show a preference in RTs for either inverted or non-inverted problems. However, when comparing their performance to that of native Hebrew speakers for aural-verbal problems presented in Hebrew, they did show the same pattern, of a preference for non-inverted over inverted problems. This preference was also apparent in the Arabic-Hebrew bilinguals' accuracy rates for aural-verbal problems presented in their two languages. Thus, they were more error prone when the structure of the addition problem mismatched the structure of number words in the language of presentation. Therefore, both the possible patterns identified in the introduction are apparent in the performance of the Arabic-Hebrew bilinguals. On the one hand, we found evidence for some adaptation to the language of presentation, mostly

in accuracy rates. On the other hand, the Arabic-Hebrew bilinguals also show evidence for enhanced flexibility, expressed as less sensitivity overall to the alignment between the order of presentation of addition problems and the structure of number words in the language.

The current results regarding the effect of order of presentation proved to be quite interesting. Previous findings comparing languages that differ in the structure of number words (Brysbaert et al., 1998; Colomé et al., 2010) support a prediction that speakers of inverted languages should prefer to solve problems that follow the order of inverted number words (unit-decade), while speakers of non-inverted languages would prefer to solve problems that follow the order of non-inverted number words (decade-unit). Colomé et al. (2010), who compared Italian and Catalan speakers, argued that language differences in the structure of number-words play a role when solving addition problems. They reached this conclusion after finding that the differences between the two groups' preference toward a particular number-word structure remained consistent both when solving problems verbally and when typing their answers on a keyboard.

Brysbaert et al. (1998), who compared Dutch and French speakers, also found that the order of presentation of addition problems, and whether it matched the structure of number words, influenced participants' performance when asked to respond verbally. Nonetheless, since these results were not replicated when participants typed their answers on a keyboard, the authors concluded that the differences between the two languages were due to a strategic adaptation to verbal output requirements instead of a direct influence of language in the addition stage.

The results of the current study show that whereas the Hebrew speakers followed the expected pattern, showing a preference for problems that follow a non-inverted order, the Arabic speakers were equally facile in responding to visual-symbolic addition problems presented in inverted and non-inverted order. In contrast, in aural-verbal presentation the Arabic-Hebrew bilinguals showed less sensitivity to order of presentation in indices of RT, but were more accurate for inverted problems in Arabic and for non-inverted problems in Hebrew. These findings suggest that the Arabic speakers are flexible and show a shift in language-order preference. In other words, it seems that by being exposed regularly to both an inverted language (Arabic) and a non-inverted language (Hebrew), they have developed the ability to process both orders equally well. It is important to note that previous studies that investigated the effect of the structure of number-words presented the experimental verbal stimuli in written form on a computer screen. Our study is the first to present participants with aurally presented addition problems without including a written representation (verbal or numeric).

Furthermore, unlike previous studies, where participants were asked to type a numerical answer or verbally answer an addition problem, the participants in the current study were asked to decide whether the problem they heard (question and answer included) was correct or incorrect. This might be an additional reason for differences found between our findings (particularly regarding the order of presentation) and those of previous studies.

In addition, the current study explored the permeability of numerical processing to influences from the L1 and the L2 in highly proficient bilinguals. This issue is closely related to the debate questioning whether conceptual representations of bilinguals are exclusively shaped by the lexical structure of L1, or whether they are open to influences from a later learned L2 (e.g., Jiang, 2002; Degani et al., 2011). The current results suggest that the numerical processing of bilinguals might be shaped by exposure to two systems differing in the structure of number words, and not exclusively determined by the L1. Further, our bilingual participants were sensitive to the language of presentation, in that they showed different preferences in the L1 and in the L2, with the latter aligning closely with the performance of native speakers of the language.

The Arabic–Hebrew speakers in the current study differed significantly in the way they processed number words in Arabic from the way they processed number words in Hebrew. They were more sensitive to unit values when they heard problems recorded in Arabic but were more sensitive to decade values when they heard similar problems recorded in Hebrew. It is true that due to our methodological decision to present problem aurally, decade identity became available earlier in Hebrew whereas in Arabic, unit identity became available first. Of course, this could have been the cause of the observed pattern of results. However, the results could also be interpreted to mean that the structure of number words in the language influences the relative emphasis of unit and decades values in arithmetic performance. In accordance with this argument, in their study, Nuerk et al. (2005) concluded

that decade distance seemed to determine responses in a number comparison task for English speakers, while overall distance was the most important predictor for German speakers, particularly when dealing with written number words.

Further Macizo et al. (2011), examined language influences by comparing Spanish/English and German/English bilinguals' performance on a number comparison task. Their results demonstrate a reverse compatibility effect observed in the L1 Spanish task for the Spanish/English bilinguals (an expected pattern for a non-inverted language), and a regular compatibility effect observed in the L1 German task for the German/English bilinguals (an expected pattern for an inverted language). However, a reverse compatibility effect was observed in the L2 English task for both groups. Since their results suggest that bilinguals process two-digit number words selectively in their L1 and L2, they concluded that bilinguals are influenced by the language of presentation and process numbers according to the structure of number words for each language. The current flexible pattern found for the Arabic–Hebrew bilinguals aligns with these results, and extends them further to aural–verbal presentation.

In summary, the use of number processing as a case study for the interactions between language and cognition in bilinguals, allowed us to clearly demonstrate two important findings: (1) the L1 does not exclusively shape the conceptual knowledge and cognitive representations, and (2) extensive exposure to an L2 can result in flexibility of representation and adaptability to different linguistic structures.

## References

- Archibald, L., Cardy, J. O., Joanisse, M. F., and Ansari, D. (2013). Language, reading, and math learning profiles in an epidemiological sample of school age children. *PLoS ONE* 8:e77463. doi: 10.1371/journal.pone.0077463
- Brybaert, M., Fias, W., and Noël, M. P. (1998). The Whorfian hypothesis and numerical cognition: is 'twenty-four' processed in the same way as 'four-and-twenty'? *Cognition* 66, 51–77. doi: 10.1016/S0010-0277(98)00066-7
- Campbell, J. I. (1994). Architectures for numerical cognition. *Cognition* 53, 1–44. doi: 10.1016/0010-0277(94)90075-2
- Colomé, À., Laka, I., and Sebastián-Gallés, N. (2010). Language effects in addition: how you say it counts. *Q. J. Exp. Psychol.* 63, 965–983. doi: 10.1080/17470210903134377
- Cook, V. (2003). "Introduction: the changing L1 in the L2 user's mind," in *Effects of the Second Language on the First*, ed. V. Cook (Clevedon: Multilingual Matters Ltd), 1–18.
- Costa, A. (2005). "Lexical access in bilingual production," in *Handbook of Bilingualism: Psycholinguistic Approaches*, eds J. F. Kroll and A. M. B. De Groot (New York: Oxford University Press).
- Degani, T., Prior, A., and Tokowicz, N. (2011). Bidirectional transfer: the effect of sharing a translation. *J. Cogn. Psychol.* 23, 18–28. doi: 10.1080/20445911.2011.445986
- Dehaene, S. (1992). Varieties of numerical abilities. *Cognition* 44, 1–42. doi: 10.1016/0010-0277(92)90049-N
- Dehaene, S., and Cohen, L. (1995). Towards an anatomical and functional model of number processing. *Math. Cogn.* 1, 83–120.
- Delazer, M., Benke, T., Trieb, T., Schocke, M., and Ischebeck, A. (2006). Isolated numerical skills in posterior cortical atrophy – an fMRI study. *Neuropsychologia* 44, 1909–1913. doi: 10.1016/j.neuropsychologia.2006.02.007
- De Smedt, B., Taylor, J., Archibald, L., and Ansari, D. (2010). How is phonological processing related to individual differences in children's arithmetic skills? *Dev. Sci.* 13, 508–520. doi: 10.1111/j.1467-7687.2009.00897.x
- Fuson, K. C., and Kwon, Y. (1992). Korean children's understanding of multidigit addition and subtraction. *Child Dev.* 63, 491–506. doi: 10.2307/1131494
- Geary, D. C., Bow-Thomas, C. C., Liu, F., and Siegler, R. S. (1996). Development of arithmetical competencies in Chinese and American children: influence of age, language, and schooling. *Child Dev.* 67, 2022–2044. doi: 10.2307/1131607
- Gelman, R., and Butterworth, B. (2005). Number and language: how are they related? *Trends Cogn. Sci.* 9, 6–10. doi: 10.1016/j.tics.2004.11.004
- Imbo, I., Vanden Bulcke, C., De Brauwier, J., and Fias, W. (2014). Sixty-four or four-and-sixty? The influence of language and working memory on children's number transcoding. *Front. Psychol.* 5:313. doi: 10.3389/fpsyg.2014.00313
- Ischebeck, A., Zamarian, L., Siedentop, C., Koppelstatter, F., Benke, T., Felber, A., et al. (2006). How specifically do we learn? Imaging the learning of multiplication and subtraction. *Neuroimage* 30, 1365–1375. doi: 10.1016/j.neuroimage.2005.11.016
- Jiang, N. (2002). Form-meaning mapping in vocabulary acquisition in a second language. *Stud. Second Lang. Acquis.* 24, 617–637. doi: 10.1017/S0272263102004047
- Jiang, N. (2004). Semantic transfer and its implications for vocabulary teaching in a second language. *Modern Lang. J.* 88, 416–432. doi: 10.1111/j.0026-7902.2004.00238.x
- Kroll, J. F., and Stewart, E. (1994). Category interference in translation and picture naming: evidence for asymmetric connections between bilingual memory representations. *J. Mem. Lang.* 33, 149–174. doi: 10.1006/jmla.1994.1008
- Laufer, B. (2003). "The influence of L2 on L1 collocational knowledge and on L1 lexical diversity in free written expression," in *Effects of the Second Language on the First*, ed. V. Cook (Clevedon: Multilingual Matters Ltd), 19–31.
- Macizo, P., and Herrera, A. (2010). Two-digit number comparison: decade-unit and unit-decade produce the same compatibility effect with number words. *Can. J. Exp. Psychol.* 64, 17–24. doi: 10.1037/a0015803

- Macizo, P., Herrera, A., Paolieri, D., and Román, P. (2010). Is there cross-language modulation when bilinguals process number words? *Appl. Psycholinguist.* 31, 651–669. doi: 10.1017/S0142716410000184
- Macizo, P., Herrera, A., Román, P., and Martín, M. C. (2011). The processing of two-digit numbers in bilinguals. *Br. J. Psychol.* 102, 464–477. doi: 10.1111/j.2044-8295.2010.02005.x
- Marian, V., Blumenfeld, H. K., and Kaushanskaya, M. (2007). The language experience and proficiency questionnaire (LEAP-Q): assessing language profiles in bilinguals and multilinguals. *J. Speech Lang. Hear. Res.* 50, 940–967. doi: 10.1044/1092-4388(2007)067
- McCloskey, M. (1992). Cognitive mechanisms in numerical processing: evidence from acquired dyscalculia. *Cognition* 44, 107–157. doi: 10.1016/0010-0277(92)90052-J
- Miller, K. F., Smith, C. M., Zhu, J., and Zhang, H. (1995). Preschool origins of cross-national differences in mathematical competence: the role of number-naming systems. *Psychol. Sci.* 6, 56–60. doi: 10.1111/j.1467-9280.1995.tb00305.x
- Miura, I. T., Kim, C. C., Chang, C. M., and Okamoto, Y. (1988). Effects of language characteristics on children's cognitive representation of number: cross-national comparisons. *Child Dev.* 59, 1445–1450. doi: 10.2307/1130659
- Nuerk, H. C., Weger, U., and Willmes, K. (2005). Language effects in magnitude comparison: small but not irrelevant. *Brain Lang.* 92, 262–277. doi: 10.1016/j.bandl.2004.06.107
- Owen, A. M., McMillan, K. M., Laird, A. R., and Bullmore, E. (2005). N-back working memory paradigm: a meta-analysis of normative functional neuroimaging studies. *Hum. Brain Mapp.* 25, 46–59. doi: 10.1002/hbm.20131
- Pesenti, M., Thioux, M., Seron, X., and Volder, A. D. (2000). Neuroanatomical substrates of Arabic number processing, numerical comparison, and simple addition: a PET study. *J. Cogn. Neurosci.* 12, 461–479. doi: 10.1162/089892900562273
- Pixner, S., Zuber, J., Hermanova, V., Kaufman, L., Nuerk, H.-C., and Moeller, K. (2011). One language, two number-word systems and many problems: numerical cognition in the Czech language. *Res. Dev. Disabil.* 32, 2683–2689. doi: 10.1016/j.ridd.2011.06.004
- Salillas, E., and Carreiras, M. (2014). Core number representations are shaped by language. *Cortex* 52, 1–11. doi: 10.1016/j.cortex.2013.12.009
- Spelke, E. S., and Tsivkin, S. (2001). Language and number: a bilingual training study. *Cognition* 78, 45–88. doi: 10.1016/S0010-0277(00)00108-6
- Unsworth, N., and Engle, R. W. (2008). Speed and accuracy of accessing information in working memory: an individual differences investigation of focus switching. *J. Exp. Psychol. Learn. Mem. Cogn.* 34, 616–630. doi: 10.1037/0278-7393.34.3.616
- Venkatraman, V., Ansari, D., and Chee, M. W. L. (2005). Neural correlates of symbolic and non-symbolic arithmetic. *Neuropsychologia* 43, 744–753. doi: 10.1016/j.neuropsychologia.2004.08.005
- Wolff, P., and Ventura, T. (2009). When Russians learn English: how the semantics of causation may change. *Biling. Lang. Cogn.* 12, 153–176. doi: 10.1017/S1366728909004040
- Zhou, X., Chen, C., Dong, Q., Zhang, H., Zhou, R., Zhao, H., et al. (2006). Event-related potentials of single-digit addition, subtraction and multiplication. *Neuropsychologia* 44, 2500–2507. doi: 10.1016/j.neuropsychologia.2006.04.003
- Zuber, J., Pixner, S., Moeller, K., and 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. *J. Exp. Child Psychol.* 102, 60–77. doi: 10.1016/j.jecp.2008.04.003

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Prior, Katz, Mahajna and Rubinsten. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Does the transparency of the counting system affect children's numerical abilities?

Ann Dowker<sup>1\*</sup> and Manon Roberts<sup>2</sup>

<sup>1</sup> Department of Experimental Psychology, University of Oxford, Oxford, UK, <sup>2</sup> Worcester College, Oxford, UK

## OPEN ACCESS

### Edited by:

Yvette Renee Harris,  
Miami University, USA

### Reviewed by:

Robert Reeve,  
University of Melbourne, Australia

Maciej Haman,  
University of Warsaw, Poland

### \*Correspondence:

Ann Dowker,  
Department of Experimental  
Psychology, University of Oxford,  
South Parks Road, Oxford OX1 3UD,  
UK  
ann.dowker@psy.ox.ac.uk

### Specialty section:

This article was submitted to  
Developmental Psychology,  
a section of the journal  
Frontiers in Psychology

**Received:** 08 December 2014

**Accepted:** 22 June 2015

**Published:** 07 July 2015

### Citation:

Dowker A and Roberts M (2015) Does the transparency of the counting system affect children's numerical abilities? *Front. Psychol.* 6:945. doi: 10.3389/fpsyg.2015.00945

The Welsh language uses a regular counting system, whereas English uses an irregular counting system, and schools within Wales teach either through the medium of Welsh or English. This provides the opportunity to compare linguistic effects on arithmetical skills in the absence of many other confounding factors that arise in international comparisons. This study investigated the hypothesis that language properties influence children's performance in certain numerical tasks by comparing the performance of 20 Welsh- and 20 English-medium Year Two pupils in non-verbal line estimations and transcoding. Groups did not differ on global arithmetic abilities, but the pupils taught through the medium of Welsh on average performed better in the non-verbal line estimation tasks than the English-medium group. This superiority was most apparent in comparisons involving numbers over 20: a result which was complicated by the fact that Welsh-medium pupils showed a lower range of error scores than the English-medium pupils. These results were thought to be related to the increased transparency of the Welsh counting system.

**Keywords:** young children, number representation, estimation, cross-linguistic, counting system, Welsh bilinguals

## Introduction

Comparisons of arithmetical performance of children in different countries have consistently shown significant differences (e.g., TIMSS, 1996; OECD, 2014). There are many possible cultural differences that may influence arithmetical development. These may include attitudes to mathematics or to academic skills in general; methods of mathematics education; the amount of time that is devoted to arithmetic teaching in school, and the economic or political situation of a country. The cultural issue that will be considered in the present article is language.

Children in Pacific Rim countries, such as China, Japan, and Korea, show superior performance in arithmetic (e.g., Fuson and Kwon, 1992; Miller et al., 1995; Miura and Okamoto, 2003), which is often attributed to the regularity of Asian counting systems. A regular counting system is purely multiplicative, meaning if the oral counting system corresponds closely to the written number word system, then children need only learn the digit names of one to nine and of multiples (i.e., 10, 100, 1000), to be able to say any number word up to 9999. Irregular counting systems on the other hand, such as English, include number words that do not show a one-to-one correspondence with the Arabic written system, such as teen numbers, e.g., "thirteen," and multiples of ten, e.g., "twenty," which consequently need to be learned separately. These differences in the degree to which the spoken and written number systems of languages coincide with each other have been suggested as an explanation for some of the results of international comparisons of arithmetical performance

of children, as a regular counting system might facilitate children's learning to count to higher numbers, and might also increase their understanding of place value concepts.

Though systematic research on the subject began relatively recently, the idea that counting systems may influence arithmetical ability has been proposed for a long time. Two-hundred years ago, Edgeworth and Edgeworth (1798) posited the English language's highly irregular counting system as a possible disadvantage to English speakers when developing arithmetical skills. Although the cross-cultural differences that have been found in arithmetical performance could be due to differences in number word systems, the fact that some countries with transparent counting systems also show superior results in mathematics performance does not prove a causal relationship. Other cross-cultural differences, such as those mentioned above, could also be contributing to mathematical performance differences. In order to assess the impact of counting system on the development of numerical skills, it is important to rule out other differences as much as possible.

Studies of children living in Wales offer a promising solution to this problem. In Wales, 80% of children receive their school instruction in English-medium schools, while 20% receive their instruction in Welsh-medium schools. All of the children are living in the same area, in similar cultural environments, and all are studying mathematics in a similar fashion according to the National Curriculum of England, Wales and Northern Ireland. Thus, it is possible to study the specific effects of language on arithmetic, independently of other educational and cultural differences. The Welsh language in fact contains both an irregular counting system and a regular counting system, but only the regular counting system is used when teaching arithmetic in schools (Roberts, 2000). For example, whereas in English, "eleven" does not correspond to the written digits, in Welsh "un deg un" is used, which translates to "one ten one." The Welsh term for "twenty-one" is "dau ddeg un" or "two tens one", which represents the tens and units more transparently than the English term.

So far, few studies have compared numerical skills in Welsh children attending English- and Welsh-medium schools. One such study was conducted by Dowker et al. (2008), in which the performance of Welsh-speaking and English-speaking children living in Wales in a variety of number processing tasks was compared. The results showed that Welsh children demonstrated superior performance in comparing two-digit numbers, but the two groups did not differ in terms of arithmetic test performance. The authors concluded that such results indicate that linguistic effects affect specific aspects of arithmetic performance as opposed to more global effects (also see Mark and Dowker, 2015). This is consistent with the fact that arithmetic performance has often been shown to be comprised of numerous components, as opposed to being a unitary function (e.g., Dowker, 2005).

Most studies of cross-linguistic differences in numerical abilities, including Dowker et al.'s earlier (2008) study of Welsh children, have looked either at possible differences in arithmetical procedures, or on transcoding and related skills such as written number comparison, or both. If counting system effects are found on arithmetic or transcoding, this may either reflect effects on

the ease of carrying out procedures in different counting systems, or differences in the internal representation of number by users of different counting systems. Relatively few studies have looked directly at whether children's internal representations of number are affected by the transparency of the counting system; and this is an important issue to resolve.

One way of studying the internal representation of number is to use the mental number line. Number line estimation tasks require participants to judge the position of a target number on a blank number line (e.g., Siegler and Booth, 2004; Booth and Siegler, 2006; Siegler et al., 2009). By calculating the discrepancy between the participant's estimate and the target number's actual location, number line estimation tasks can be used to assess participants' internal mental number line representation. Research has suggested that overall number sense revolves around a mental number line, and that children's mental number line representations correlate with and are casually linked to arithmetic performance (e.g., Siegler and Booth, 2004; Booth and Siegler, 2006, 2008; Link et al., 2014). Furthermore, a developmental shift has been consistently observed between the ages of 5 and 8 years, in which children rely increasingly on linear representations of numerical magnitudes as opposed to a reliance on logarithmic representations, with the shift being age dependent, and occurring for smaller scales before a linear representation of larger scales emerges (Booth and Siegler, 2008). Using non-verbal number line estimation tasks, in which the target number is presented visually rather than verbally, is particularly advantageous as makes it possible to assess linguistic effects on mental number line representations without the confounding factor of verbal comprehension, which might be separately influenced by language.

The few studies that have so far investigated the mental number line representations of children from different linguistic backgrounds have given somewhat conflicting results, though most have suggested at least some linguistic influence. Siegler and Mu (2008) found that Chinese kindergarten children performed better than their American counterparts on mental number line estimation tasks involving a number line spanning from 1 to 100. Laski and Yu (2014) found that both Chinese and Chinese-American children performed better than monolingual English-speaking American children, though they also found that children in China performed better on these tasks than Chinese-American children, suggesting that educational factors were more important than linguistic factors. By contrast, Muldoon et al. (2011) did not find such a difference between Chinese and Scottish 4- and 5-year-olds; and indeed when smaller number lines from 0 to 10 and 0 to 20 were included, the Scottish children performed better.

Helmreich et al. (2011) looked at another language group, with a counting system that is *less* transparent than most European counting systems. German speakers have a counting system that has the potentially confusing property of inversion of tens and units: e.g., the number that is written as 24 is spoken as "vier und zwanzig" (twenty-four). Helmreich et al compared German- and Italian-speaking children's estimation accuracy in a non-verbal number line estimation task in which the number line spanned from "0" to "100." The German children did indeed

perform significantly worse than the Italian-speaking children on the mental number line task, though not in tests of global arithmetical ability.

The main aim of the current study was to extend the findings of Dowker et al. (2008) to investigate whether Welsh and English medium children in Wales would differ in the precision of their non-verbal number line estimation. In other words, does the language of school instruction affect the mental number line in children who are otherwise having similar cultural and educational experiences? Number line estimation tasks do not form part of the teaching syllabus in Wales, meaning children's performance in this task should be largely immune to teaching effects. Additionally, we wished to assess and compare the children's general arithmetical ability and their numerical transcoding abilities: their ability to read and write two-digit numbers.

The performance of 20 Welsh-medium and 20 English-medium children was compared regarding the British Abilities Scale (BAS) Number Skills test, writing to dictation, reading numbers aloud, and a non-verbal line estimation task, which included number lines from 0–20 and 0–100.

We predicted that the Welsh-medium children would score higher on the number line estimation tasks, and that this would be especially true of the number line from 1 to 100, as this places greater demands on the ability to represent multi-digit numbers, for which a transparent counting system would provide an advantage. On the basis of Dowker et al.'s (2008) findings about Welsh-medium children's better performance on two-digit number comparison, we also predicted that the Welsh-medium children would perform better on the tasks involving reading and writing numbers. However, we predicted that there would be no difference between the two groups in BAS Number Skills test performance, based on Dowker et al.'s (2008) earlier findings and their suggestion that linguistic effects on mathematics may be specific rather than global.

## Method

### Ethical Approval

Ethical approval for this study was obtained from Oxford University's Central University Research Ethics Committee.

### Participants

Forty children, drawn from the Year Two classes of two state primary schools in Cardiff and one state primary school from the Rhondda Cynon Taf area, took part in the study. The data from all participants were included in the analysis. Written consent was obtained from all parents or guardians. One Cardiff-based school taught through the medium of Welsh, from which 20 children (10 girls) took part. The other two schools were English-medium schools, and 10 participants from each school (20 in all, including 14 girls) took part. All the children in the Welsh-medium schools were taught exclusively through the medium of Welsh, but 13 of the 20 children spoke English as a first language. Though taught through different languages, mathematics teaching followed exactly the same curriculum in the three schools. The 20 Welsh-medium children were

compared with the 20 English-medium children. All were tested at the same time of their school year, but they turned out to be somewhat different in age. The mean age of the Welsh-medium school pupils was 6 years and 5 months ( $SD = 0.30$ ; range 73–85), and the English-medium school pupils 6 years and 7 months ( $SD = 0.35$ ; range 73–84). The age difference between the two groups was significant [ $t_{(38)} = 2.38, p = 0.022, d = 0.75$ ]. All children had normal or normal-to-corrected vision.

### Tasks and Procedure

The children completed four tasks: the BAS Number Skills test, which is a standardized test that assesses written calculation (Elliott et al., 1997), two transcoding tasks (writing to dictation and reading aloud), and a non-verbal number line estimation task.

In the *writing Arabic numbers to dictation task*, participants were required to write down 32 different Arabic numbers (two single-digit, 10 double-digit, and 20 3-digit numbers) that were presented verbally one by one by the experimenter.

In the *reading Arabic numbers aloud task*, participants were required to read aloud 32 different Arabic numerals (two single-digit, 10 double-digit, and 20 three-digit numbers) that were presented on a computer screen one by one. In both transcoding tasks, items were scored with a 0 for every incorrect answer, and 1 for every correct answer.

The *number line task* was a pen and paper task that required participants to estimate the position of a visually presented number (as opposed to verbally presented) on an empty number line, without counting or using any other strategy other than estimation. The number lines were 10 cm long, and labeled with "0" on the left end, and "20" or "100" on the right end. Each number to be estimated was presented centrally above each empty individual number line in Arabic notation. Participants estimated the position of the numbers 12, 1, 13, 4, 15, 19, 7, 17, and 5 on 0–20 number lines, and the position of the numbers 27, 2, 64, 35, 7, 13, 99, 75, 47, 3, 11, 82, 95, 9, 17, 6, 18, and 53 on 0–100 number lines. Before beginning the task, participants were presented with an orienting problem for each of the two different number lines, where they were required to estimate the position of 10 on the 0–20 number line, and 50 on the 0–100 number line for practice purposes.

The children were tested in one-to-one single sessions with the experimenter. The tasks were explained and conducted in Welsh for the Welsh-medium education children and in English for the English-medium education children. Each trial was presented sequentially, and no feedback on performance was provided for any of the trials, including the practice trials.

The data were analyzed using IBM SPSS 20.

## Results

### Overall Mean Scores

The mean raw score on the British Abilities Scales Basic Number Skills test (henceforward referred to as BAS) was 8.25 (s.d. 3.24) and the mean standard score was 107.98 (s.d. 11.9). The mean score for the Reading Aloud test was 20.98 (s.d. 8.42) and the Writing test was 19.75 (s.d. 7.95).

To obtain number line estimation scores, the distance between the true position of the number that was presented and the position of the number corresponding to the child's estimate on the number line was measured to the nearest millimeter. These deviation measures were then averaged for each participant individually to give two mean estimation error scores; one score for the 0–20 number line estimations, and one for the 0–100 number line estimations. These were also averaged to obtain an overall mean estimation error for each participant.

The mean estimation error score for the 0–20 number line was 12.12 mm. (s.d. 9.43) and the mean estimation error for the 0–100 number line was 20.36 mm. (s.d. 7.15). The overall mean estimation error score was 17.58 (s.d. 6.98).

### Correlations Between Age, BAS Scores and Other Measures

A correlation table is given in the Supplementary Material.

Pearson product-moment correlations were carried out between Age in months and the other measures. Age did not correlate significantly with the BAS raw score or standard score, nor with the estimation error measures, though there was a trend toward a significant negative correlation with errors for the 0–100 number line, i.e., for older children to perform slightly better [ $r_{(38)} = -0.295$ ;  $p = 0.065$ ]; but it did correlate very significantly with Reading Aloud [ $r_{(38)} = 0.53$ ;  $p < 0.001$ ] and Writing [ $r_{(38)} = 0.46$ ;  $p = 0.003$ ].

Pearson product-moment correlations were carried out between BAS raw score and the other measures. The BAS raw score showed a significant negative correlation with estimation errors overall [ $r_{(38)} = -0.439$ ;  $p = 0.005$ ] and with errors for the 0–100 number line [ $r_{(38)} = -0.479$ ;  $p = 0.002$ ], though it did not correlate significantly with errors for the 0–20 number line [ $r_{(38)} = -0.25$ ;  $p = 0.12$ ]. It also correlated significantly with Reading Aloud [ $r_{(38)} = 0.69$ ;  $p < 0.001$ ] and Writing [ $r_{(38)} = 0.73$ ;  $p < 0.001$ ]. Correlations between the BAS standard score and the other measures were very similar to those between the BAS raw score and the other measures.

The overall estimation error score showed a significant negative correlation with Reading Aloud [ $r_{(38)} = -0.6$ ;  $p < 0.001$ ] and Writing [ $r_{(38)} = 0.49$ ;  $p < 0.001$ ].

For a more detailed list of correlations, see the Supplementary Material.

### Comparison Between Groups: Analyses of Covariance

As the language groups differed significantly in Age, and as Age correlated significantly with some measures, Analyses of Covariance were carried out with Age as a covariate.

To compare the Reading Aloud and Writing scores of the Welsh- and English-medium groups, two univariate ANCOVAs were conducted with language as the fixed factor, age as a covariate, and the Reading and Writing scores as dependent variables. Group differences did not approach significance for either task.

To compare the estimation errors of the Welsh- and English-medium groups, three univariate ANCOVAs were conducted with language as the fixed factor, age as a covariate, and the three

different mean estimation error scores as dependent variables. The ANCOVA revealed no significant group difference for the 0–20 number line estimation errors [ $F_{(1, 37)} = 2.77$ ;  $p = 0.11$ ; partial  $\eta^2 = 0.05$ ], but did reveal a significant difference for the overall estimation error score [ $F_{(1, 37)} = 4.36$ ,  $p = 0.044$ ; partial  $\eta^2 = 0.11$ ], and a borderline significant difference for estimation errors in the 0–100 number line task, [ $F_{(1, 37)} = 3.77$ ,  $p = 0.06$ ; partial  $\eta^2 = 0.092$ ]. Estimation errors were lower in the Welsh-medium group for all these tasks, though the difference only approached significance for the 0–100 task ( $M = 19.05$ ,  $SD = 5.52$  for the Welsh-medium pupils compared to  $M = 21.69$ ,  $SD = 8.42$  for the English-medium pupils); and reached it for the overall estimation error score ( $M = 16.06$ ,  $SD = 4.66$  for the Welsh-medium pupils and  $M = 19.1$ ,  $SD = 8.57$  for the English-medium pupils).

### Further Analyses of Estimation Error Scores

As the 1–100 line included both comparisons of numbers over 20 and numbers under 20, analyses were carried out on mean errors for both types of number separately, to elucidate whether group differences related to the entire number line, or just to the larger numbers. For numbers under 20, there was no significant group difference at all [ $F_{(1, 37)} = 0.844$ ,  $p = 0.89$ ; partial  $\eta^2 = 0.001$ ]. However, for numbers over 20, the group difference was significant [ $F_{(1, 37)} = 9.14$ ,  $p = 0.003$ ; partial  $\eta^2 = 0.274$ ]. The Welsh-medium pupils had a mean estimation error score of 14.67 ( $SD = 4.88$ ) as compared with 21.09 ( $SD = 13.5$ ) for the English-medium pupils.

Since the standard deviation was higher in the English than the Welsh group for numbers over 20 on the 1–100 line, analysis of variance may not be a fully adequate measure; and non-parametric analyses were also carried out. A Kolmogorov-Smirnov independent-samples test failed to reach significance ( $p = 0.172$ ); while a Moses Test of Extreme Reaction showed a significant difference between the ranges of the two groups, with a higher range in the English group ( $p = 0.02$ ).

For the 00–20 number line, analyses were carried out on mean errors for teen numbers vs. numbers below 10. To compare the estimation errors of the Welsh- and English-medium groups, two univariate ANCOVAs were conducted with language as the fixed factor, age as a covariate, and the two different mean estimation error scores (for numbers under 10 and over 10 on the 1–20 number line) as dependent variables. The ANCOVA revealed no group difference even approaching significance for numbers under 10 [ $F_{(1, 37)} = 0.005$ ;  $p = 0.95$ ], but did reveal a significant difference for numbers over 10 [ $F_{(1, 37)} = 5.43$ ,  $p = 0.025$ ; partial  $\eta^2 = 0.13$ ]. For the numbers over 10, Welsh-medium children made lower estimation errors ( $M = 11.64$ ;  $SD = 10.39$  for the Welsh-medium group and  $M = 18.76$ ;  $SD = 13.62$  for the English-medium group).

### Discussion

This study aimed to extend Dowker et al.'s (2008) study of Welsh children, and to examine the role of language in children's transcoding skills and non-verbal number line estimations by

comparing two groups of school pupils for which cultural and educational variables were not strong confounding variables.

The results suggest that the transparency of the counting system may indeed have an influence on children's number representation, far more than on arithmetical skills. There was no evidence in this study for better arithmetic or transcoding in the Welsh-medium children. The two groups performed similarly in the BAS standardized test, and in the reading and writing of numbers.

However, after controlling for age, the Welsh-medium children performed better on the number line estimation task. This is congruent with the findings of Siegler and Mu (2008), Laski and Yu (2014), and Helmreich et al. (2011), indicating that the transparency of the counting system may influence not only the ease of carrying out procedures with numbers (although in the present study there was no evidence for this), but their representations of numbers.

This is particularly striking, because in this study there were no apparent environmental differences between the groups, other than the linguistic ones. The children attended apparently similar schools, and were all studying mathematics according to the same curriculum. It should be noted that the majority of the Welsh-medium children spoke English at home, which makes any effect of school instruction medium even more striking. While it is not possible completely to rule out the influence of some non-linguistic characteristics of a particular school or teacher, this seems unlikely, especially as the groups did not differ on the standardized arithmetic test.

The group differences varied both with the nature and extent of the number line and with the size of the numbers involved in particular tasks. The Welsh-medium group made lower estimation errors, and this difference was significant overall and approached significance for the 0–100 number line, but not for the 0–20 number line. This is consistent with the prediction that language would be likely to affect number line estimation more for multi-digit numbers. This prediction was indeed supported by further analyses of estimates for different sizes of numbers within the 0–20 and 0–100 number lines. The groups differed on the 0–20 number line for their estimates of teen numbers, but not of numbers under 10. On the 0–100 number line, they differed significantly for numbers over 20, but not at all for numbers from 0 to 20 (single digit or teen). Thus, whether the groups differed for teen numbers or not seemed to differ according to the surrounding numerical context, but there is more evidence that the groups differed for numbers over 20, and did not differ for numbers under 10. In other words, the precision of numerical representations was most affected by language for numbers that required an understanding of the relationship between tens and units, which appears to be facilitated by a transparent counting system. These representations were also those most related to arithmetical performance: the BAS Arithmetic test correlated significantly with the error score on the 1–100 number line but not the 0–20 number line.

There is, however, an important qualification with regard to the statement that the groups differed significantly for numbers over 20. The *range* of error scores was significantly higher for the English medium pupils. Presumably as a result of this, a

non-parametric comparison between the language groups for this group of numbers failed to reach significance. The greater range of scores for the English medium group is intriguing in itself. One possible explanation may be that the transparent counting system of Welsh constrains the strategies for estimation, while the less transparent English system provides fewer constraints and cues, leading to greater variability in strategies and thereby in scores. A yet more interesting possible explanation is that the transparent counting system constrains representation as such, and that this is far more variable in a more opaque counting system. Clearly, more research needs to be done, involving a wider variety of representational tasks and a larger set of transparent and opaque counting systems.

The hypothesis that reading and writing two-digit numbers would be affected by the language group was not supported at all. It may be that these skills, which are school-taught, depend more on specific teaching, which would be similar in the Welsh- and English-medium schools, and do not depend strongly on internal numerical representations. Although this finding might seem to conflict with the earlier results of Dowker et al. (2008) that indicated that Welsh-medium children were better than English-medium children at dealing with two-digit numbers, the tasks were rather different. The task in Dowker et al.'s (2008) study involved reading *and comparing* two-digit numbers; and it may be that the comparison element of the task was more dependent on internal representations.

Intriguingly, what did correlate quite strongly with the number reading and writing tasks was chronological age, even though the age range in this study was restricted and the children were all tested at the same time in their school year. Further research needs to be done to see whether maturation or perhaps some specific aspect of home experience is particularly important in the development of these skills. The finding that young children can show a significant age correlation, even within a quite restricted age range, with some numerical abilities but not others is congruent with earlier studies with somewhat younger children (Dowker, 2008), and clearly needs further exploration.

Thus, the results imply that greater transparency in a language's counting system may lead to the developmental shift in children's mental number line representation (that is, the shift from a logarithmic representation to a linear representation) occurring at an earlier age.

Given that previous research has documented this representation's link with overall arithmetic ability, the advantage of being taught arithmetic through the medium of a language that utilizes a regular counting system may be more widespread than previously thought. It would be interesting to carry out longitudinal studies to investigate whether this advantage persists longitudinally, and whether it predicts any other aspects of later arithmetic. The Welsh-medium children did better at the number line estimation tasks; these tasks correlated with a standardized arithmetic test; but the Welsh-medium children did not at this stage do better at the standardized arithmetic test. Might the number line advantage correlate with standardized arithmetic test differences, or with differences in more specific aspects or arithmetic later on? Cross-sectional studies of a wider variety of age-groups would also yield interesting information

about how language effects on numerical skills might change with age.

Though the results do suggest that the transparency of a language's counting system has an effect on some aspects of number processing, some caution is needed to drawing extremely strong conclusions, as the groups, though from similar backgrounds, were not matched in advance on all possible factors, and in particular there was an unexpected small but significant difference in age. This was controlled for in the analyses by including age as a covariate, as age did show a surprising level of correlation with some measures; but this is not a completely satisfactory solution, and future studies should ensure matching for age.

Moreover, direct measures of proficiency in the two languages were not obtained. Although the English-medium children studied Welsh formally as a second language at school, they may not have been bilingual in the same sense as children who speak English at home and Welsh at school. At present, we cannot be certain about the extent to which group differences are the result of general proficiency in Welsh; Welsh medium mathematics instruction in particular; or the experience of bilingualism. It would be desirable for future studies to explore the issue further. In particular, it would also be interesting to study Welsh-English balanced bilinguals, and to investigate whether different results would be obtained when testing the same children in Welsh vs. English.

These results may have some implications for educational practice. As stated in the Introduction, there is evidence that the development of linear representations of numerical magnitude contributes to arithmetical development (Booth and Siegler,

2008). The present study also supports a few earlier studies in suggesting that this development may be facilitated by regular counting systems and impeded by irregular counting systems. Perhaps English-medium schools, or any schools that teach through the medium of a language that has an irregular counting system, could investigate ways of helping children generate linear representations of number. For example, one way of achieving this might be through the playing of board games in which the counters are moved linearly across equidistant spaces, thus providing a transparent representation of numerical magnitude (Siegler and Ramani, 2009).

The main conclusion of this study is that the regularity and transparency of the Welsh counting system may help children not only in learning the correspondence between written and oral representations of number but also in the development of non-verbal numerical magnitude representations. Therefore, the influence of language should be considered when teaching number processing, especially when teaching children who struggle with mathematics.

## Acknowledgments

We would like to thank the teachers and children in the three schools that took part in the study.

## Supplementary Material

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpsyg.2015.00945>

## References

- Booth, J. L., and Siegler, R. S. (2006). Developmental and individual differences in pure numerical estimation. *Dev. Psychol.* 42, 189–201. doi: 10.1037/0012-1649.41.6.189
- Booth, J. L., and Siegler, R. S. (2008). Numerical magnitude representations influence arithmetic learning. *Child Dev.* 79, 1016–1031. doi: 10.1111/j.1467-8624.2008.01173.x
- Dowker, A. (2005). *Individual Differences in Arithmetic: Implications for Psychology, Neuroscience and Education*. Hove: Psychology Press.
- Dowker, A. (2008). Individual differences in numerical abilities in preschoolers. *Dev. Sci.* 11, 650–654. doi: 10.1111/j.1467-7687.2008.00713.x
- Dowker, A., Bala, S., and Lloyd, D. (2008). Linguistics influence of mathematical development: how important is the transparency of the counting system? *Philos. Psychol.* 21, 523–538. doi: 10.1080/09515080802285511
- Edgeworth, M. and Edgeworth, R. L. (1798). *Practical Education*. London: J. Johnson.
- Elliott, G., Murray, D., and Pearson, L. (1997). *British Abilities Scales, 2nd Edn*. Windsor: National Foundation for Educational Research.
- Fuson, K., and Kwon, Y. (1992). Korean children's understanding of multi-digit addition and subtraction. *Child Dev.* 63, 491–506. doi: 10.2307/1131494
- Helmreich, I., Zuber, J., Pixner, S., Kaufmann, L., Nuerk, H.-C., and Moeller, K. (2011). 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. *J. Cross Cult. Psychol.* 42, 598–613. doi: 10.1177/0022022111406026
- Laski, E. V., and Yu, Q. Y. (2014). Number line estimation and mental addition: examining the potential roles of language and education. *J. Exp. Child Psychol.* 117, 29–44. doi: 10.1016/j.jecp.2013.08.007
- Link, T., Huber, S., Nuerk, H. C., and Moeller, K. (2014). Unbounding the mental number line—new evidence on children's spatial representation of numbers. *Front. Psychol.* 4:1021. doi: 10.3389/fpsyg.2013.01021
- Mark, W., and Dowker, A. (2015). Linguistic influence on mathematical development is specific rather than pervasive: revisiting the Chinese Number Advantage in Chinese and English children. *Front. Psychol.* 6:203 doi: 10.3389/fpsyg.2015.00203
- Miller, K. F., Smith, C. M., Zhu, J., and Zhang, H. (1995). Preschool origins of cross-national differences in mathematical competence: the role of number-naming systems. *Psychol. Sci.* 6, 56–60. doi: 10.1111/j.1467-9280.1995.tb00305.x
- Miura, I., and Okamoto, Y. (2003). "Language supports for mathematics understanding and performance," in *The Development of Arithmetical Concepts and Skills*, eds A. Baroody and A. Dowker (Mahwah, NJ: Lawrence Erlbaum Associates, Inc.), 229–242.
- Muldoon, K., Simms, V., Towse, J., Menzies, V., and Yue, G. (2011). Cross-cultural comparisons of 5-year-olds' estimating and mathematical ability. *J. Cross Cult. Psychol.* 42, 669–681. doi: 10.1177/0022022111406035
- OECD. (2014). *PISA 2012 Results: What Students Know and Can Do – Student Performance in Mathematics, Reading and Science (Volume I, Revised Edition, February 2014)*. Paris: PISA OECD Publishing.
- Roberts, G. (2000). Bilingualism and number in Wales. *Int. J. Biling. Educ. Biling.* 3, 44–56. doi: 10.1080/13670050008667699
- Siegler, R. S., and Booth, J. L. (2004). Development of numerical estimation in young children. *Child Dev.* 75, 428–444. doi: 10.1111/j.1467-8624.2004.00684.x
- Siegler, R. S., and Mu, Y. (2008). Chinese children excel on novel mathematical problems even before elementary school. *Psychol. Sci.* 19, 759–763. doi: 10.1111/j.1467-9280.2008.02153.x

- Siegler, R. S., and Ramani, G. B. (2009). Playing linear board games—but not circular ones—improves low-income preschoolers' numerical understanding. *J. Educ. Psychol.* 101, 545–560. doi: 10.1037/a0014239
- Siegler, R. S., Thompson, C. A., and Opfer, J. E. (2009). The logarithmic- to-linear shift: one learning sequence, many tasks, many time scales. *Mind Brain Educ.* 3, 143–150. doi: 10.1111/j.1751-228X.2009.01064.x
- TIMSS (Trends in International Mathematics and Science Study). (1996). *Highlights of Results from TIMSS: Third International Mathematics and Science Study*. Chestnut Hill, MA: TIMSS International Study Centre.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

*Copyright © 2015 Dowker and Roberts. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*



# Linguistic influence on mathematical development is specific rather than pervasive: revisiting the Chinese Number Advantage in Chinese and English children

Winifred Mark<sup>1\*</sup> and Ann Dowker<sup>2</sup>

<sup>1</sup> Department of Psychology, University of Hong Kong, Hong Kong, China

<sup>2</sup> Department of Experimental Psychology, University of Oxford, Oxford, UK

## Edited by:

Yvette Renee Harris, Miami University, USA

## Reviewed by:

Anna M. Borghi, University of Bologna, Italy

Koen Luwel, KU Leuven, Belgium

Barbara W. Sarnecka, University of California, Irvine, USA

## \*Correspondence:

Winifred Mark, Department of Psychology, University of Hong Kong, 6/F, The Jockey Club Tower, Centennial Campus, Pok Fu Lam Road, Hong Kong, China  
e-mail: wmark@hku.hk

The relative linguistic transparency of the Asian counting system has been used to explain Asian students' relative superiority in cross-cultural comparisons of mathematics achievement. To test the validity and extent of linguistic transparency in accounting for mathematical abilities, this study tested Chinese and British primary school children. Children in Hong Kong can learn mathematics using languages with both regular (Chinese) and irregular (English) counting systems, depending on their schools' medium of instruction. This makes it possible to compare groups with varying levels of exposure to the regular and irregular number systems within the same educational system, curriculum, and cultural environment. The study included three groups of first/second graders and third/fourth graders with varying degrees of experience to the Chinese language and counting systems: no experience (UK;  $n = 49$ ); spoke Chinese at home and learnt to count in English at school (HK-E;  $n = 43$ ); spoke Chinese at home and learnt to count in Chinese at school (HK-C;  $n = 47$ ). They were compared on counting, numerical abilities and place value representation. The present study also measured nonverbal reasoning, attitude toward mathematics, involvement of parents, and extra-curricular mathematics lessons to explore alternative explanations of children's numeric ability. Results indicated that students in HK-C were better at counting backward and on the numeric skills test than those in HK-E, who were in turn better than the UK students. However, there was no statistical difference in counting forward, place value understanding, and a measure of arithmetic. Our findings add to existent literature suggesting that linguistic transparency does not have an all-pervasive influence on cross-national differences in arithmetic performance.

**Keywords:** linguistic transparency, counting system, arithmetic, cross-cultural, Chinese Number Advantage

## INTRODUCTION

International comparisons of children's arithmetic performance, such as the Trends in International Mathematics and Science Study, consistently showed that Asian students outperformed their Western counterparts (Stedman, 1997; Mullis et al., 2000, 2008; Provasnik et al., 2012). While many individual and sociological factors could influence mathematics learning, the current study focused on linguistic influences on early mathematics learning. Recent years have seen a surge in empirical literature on the role of language in accounting for cross-cultural disparities in children's number understanding and arithmetic competence (Fuson and Kwon, 1992; Aunio et al., 2004, 2006, 2008; Cheng and Chan, 2005; Rasmussen et al., 2006; Wang et al., 2008; Göbel et al., 2011; Krinzinger et al., 2011; Pixner et al., 2011; Zhao and Singh, 2011; Klein et al., 2013; Cankaya et al., 2014). Linguistic influences on mathematics learning warrant interest because the capacity to name and manipulate numeric quantities has been used to explain why human mathematical abilities could develop beyond the rudimentary number sense observed in animals (Dehaene, 1997). If the representation for large quantities and algorithms for calculation were underlay by language, it follows that distinct linguistic

characteristics could lead to differential computational efficiency and arithmetic understanding.

There is a lot of debate about the extent to which language affects thought in general; but some evidence suggests that abstract concepts are more influenced than concrete ones by linguistic diversity (Gentner and Boroditsky, 2001; Borghi et al., 2011; Borghi and Binkofski, 2014). As number is a highly abstract concept, one might expect it to be more influenced by linguistic diversity than some other domains.

One linguistic characteristic that could influence children's mathematics learning is the way in which numbers and arithmetical relationships are expressed in the counting system. It has been suggested that the superior arithmetic performance of Chinese and other Asian students could be explained by the relative *linguistic transparency* of many Asian counting systems (Fuson and Kwon, 1991; Miller et al., 2005; Ng and Rao, 2010), termed the 'Chinese Number Advantage' (CNA). Transparent number systems give a clear and consistent representation of the base system (base-ten in most languages). One example is the Chinese counting system, where the boundary between 10 and 11 is explicit in both written and spoken forms. The Chinese word for 11 is 十一 (*shi yi*), literally

'ten-one'; that for 12 is 十二 (*shi er*), literally 'ten-two,' and so on. The same rule applies for larger numbers, such that 20 is 二十 (*er shi*) 'two-ten,' 59 is 五十九 (*wu shi jiu*) 'five-ten-nine' and so on. Hence, new numbers could easily be inferred in Chinese, and it is clear that the numbers are organized according to a base-ten system.

Edgeworth and Edgeworth (1798) suggested more than 200 years ago that English-speakers might be at a disadvantage compared with speakers of other languages due to the relatively irregular English counting system. This gained empirical support from Miller et al. (1995), who found that Chinese and American 4- and 5- year olds performed similarly in learning to count up to 12, but the Chinese students were about a year ahead of the American children in the further development and counting of higher numbers. In contrast to regular counting systems, the English words *eleven* and *twelve* do not provide clear clues for their cardinality nor the base system. Those well-versed in the history of numbers might recognize that the English words for 11 and 12 reflected historical relations to the Old Saxon words *ellevan* and *twelif*, literally 'one-left' and 'two-left' respectively after 10 has been subtracted. However, this information is not apparent to young learners! In addition, various phonemic modifications further complicate number learning for English children: In 13–19, *ten* becomes *-teen*, *three* becomes *thir-*, and *five* becomes *fif-*. Above 19, *ten* becomes *-ty* for tens starting from 20, *two* becomes *twen-* in the twenties and *four* becomes *for-* in the forties.

English children also had more difficulties than speakers of some other languages in acquiring the base-ten system. Since English children must learn *one* through *twelve* by rote learning, the base-ten system might be scaffolded. Experimental evidence was provided by cross-cultural studies on six-year-olds using regular counting systems such as Chinese, Japanese, and Korean versus children from less regular counting systems such as French, Swedish, and the U. States. (Miura et al., 1988, 1993; Miura and Okamoto, 2003). Children were asked to represent numbers with cubes representing single units and ten-segmented blocks representing tens. It was found that children from regular counting systems were more likely to use bars and cubes in combination to represent numbers, while children from less regular counting systems were more likely to count out the exact number of cubes. Failure to take advantage of tens-bars suggested poorer understanding of the base-ten system.

The greater transparency of base system might make place value easier to grasp in a regular counting system (Miura and Okamoto, 2003). Place-value knowledge refers to the knowledge of the value of each digit by considering its place in a multi-digit number, such that each '5' in 555 is understood as 5 hundreds, 5 tens, and 5 units, respectively. Such knowledge is essential for arithmetic computations. The regular Chinese number system can be directly mapped onto Arabic numbers; for example, 17 is 'ten-seven' in Chinese, making it obvious that the '1' is a '10.' In contrast, place values of English numbers are obscured by the three forms of ten (*ten*, *-teen*, and *-ty*), and the fact that the order of reading numbers does not necessarily align with the Arabic numbers (e.g., *seventeen* vs. *seventy*). Such irregularities mask place values and hinder English children's arithmetic development.

Despite the linguistic advantages that the Chinese number system potentially afforded, some considered that the CNA could not be an adequate explanation for Asian children's superiority over Western children in nearly *all* mathematical domains (Ackerman, 1988). The many other cultural differences between Asian and Western children, such as quantity and quality of mathematics teaching (Saxton and Towse, 1998), attitudes of parents and personal motivation toward mathematics (Stevenson et al., 1993) weaken the CNA. Research conducted in Wales (MacLean and Whitburn, 1996; Dowker and Lloyd, 2005; Dowker et al., 2008) offered important insights in this regard, since groups with varied levels of exposure to regular (Welsh) and irregular (English) number systems could be compared. Dowker et al. (2008) found that Welsh children were facilitated on reading and comparing two-digit numbers, but not on all arithmetic tests. They concluded that linguistic transparency could not on its own explain the cross-national differences in arithmetic, thus providing indirect evidence against the CNA.

In a similar attempt to distinguish language and cultural effects, and to test the CNA directly, the present study recruited British and Hong Kong primary school students. The Hong Kong educational system is based upon the British system, reflecting its history as a British colony. Mathematics could be taught in a regular (Chinese) or irregular (English) counting system, depending on the medium of instruction of the school. It is hence possible to compare the mathematical performance of children who received either English- or Chinese-medium schooling, within the same educational system, curriculum, and cultural environment. Our study adds to the literature in that it is one of the first studies to take advantage of the Chinese/English medium of instruction system in Hong Kong to study linguistic influences in Mathematics. Our study also attempts to extend Dowker et al.'s (2008) Welsh study, as it also compares groups of children taught in different languages within otherwise similar settings. Furthermore, this study serves as a supplement to existing CNA studies, many of which compare Chinese and Finnish (Aunio et al., 2004, 2006, 2008).

Three groups of primary school children with varying degrees of experience with the Chinese language and counting system were compared in this study—those who had no experience (British students); those who spoke Chinese at home but learnt Mathematics in English (students in English-medium schools in Hong Kong); and those who spoke and learnt Mathematics in Chinese at both home and school (students in Chinese-medium schools in Hong Kong). The English- and Chinese-medium schoolchildren in Hong Kong differed mainly in terms of the linguistic medium used in their school instruction, but otherwise had similar cultural and educational experiences; while the British children were of course growing up within a different culture and educational system. They were all given a non-verbal intelligence measure, a test of numerical skills, a test of place value representations, and an attitude toward mathematics questionnaire. As both Chinese- and English-medium schools in Hong Kong followed the same mathematics curriculum, the two groups of Hong Kong children differed primarily in the language in which they learnt mathematics. Testing Hong Kong students taught in different media of instruction allowed us to tease apart whether it is the exposure to

the Chinese language *per se* or the use of the Chinese counting system that influenced mathematical ability. British students served as a control group for exposure to the Chinese language, while students in the English-medium school in Hong Kong served as the control group for formal instruction of the Chinese counting system. The present study also took into account the role of children's attitude toward mathematics and involvement of parents, both of which were often omitted in previous cross-linguistic studies (MacLean and Whitburn, 1996; Dowker et al., 2008).

Based on the CNA, it was hypothesized that (1) students in Hong Kong would perform better than British students on all numerical tasks, including counting, place value knowledge and the numerical skills test; (2) within Hong Kong, students in Hong Kong Chinese-medium schools would perform better than those in English-medium schools. In order to study the impact of duration in use of Chinese number system in numerical skills, we recruited a younger group (first-/second-graders) and an older group of children (third-/fourth-graders).

## MATERIALS AND METHODS

### PARTICIPANTS

A total of 159 children from two primary schools in Hong Kong and two primary schools in Oxford, UK participated in the testing. As the proximity of primary schools to students' homes constitutes a major factor in primary school enrollment, the socioeconomic status (SES) of the catchment area in which the schools were situated could be considered proxy to the SES of their students. In this regard, the schools in Hong Kong and UK were located in predominantly middle-class areas. Testing in Hong Kong was done in August while testing in the UK was done in October of the same year. To ensure similarity of age and years in school, Hong Kong students were at the end of their first and third grade while UK students were at the start of their second and fourth grade. Written informed consent was obtained from parents of all participants. The study was approved by the Central University Research Ethics Committee of University of Oxford.

At the Chinese-medium school in Hong Kong (henceforth HK-C), Cantonese was the first language for all children, who came from Chinese-speaking homes. They received a Chinese-medium education, and were taught Mathematics in Cantonese. There were 25 first-graders and 25 third-graders from HK-C. At the English-medium school in Hong Kong (henceforth HK-E), Cantonese was the first language of the children, and they spoke Chinese at home. However, they received education in English for most school subjects including Mathematics. There were 37 first-graders and 16 third-graders from HK-E. At the British school in Oxford (henceforth UK), English was the first language of the children. They spoke English at home and at school, with no exposure to the Chinese language. There were 26 second-graders and 30 fourth-graders from UK.

### MEASURES

Measures employed were translated and back-translated from the English-version into Chinese by the first author and a bilingual experienced mathematics teacher, respectively. Two experienced

mathematics teachers at a Chinese-medium primary school then reviewed all items.

### Demographic and background information

Participants were asked about their age, grade and whether they attended kindergarten. To investigate the effect of additional mathematical instruction and parental involvement, participants were asked whether they attended mathematical classes outside of school and whether their parents helped them with their homework in general, as well as in math homework in particular.

### Counting

Participants counted aloud from 1 to 30 and then backward from 30 to 1. Hesitations (more than 3 s delay), missing numbers, and incorrect sequence were recorded.

### Numerical abilities

All children completed the British Abilities Scales (BAS) Basic Number Skills test, which involved recognizing and reading two-/three-digit numbers, as well as solving simple written calculations. Scores in *addition*, *subtraction*, *multiplication*, *division*, *fraction*, and *decimals* were added to compute a 'purer' measure of arithmetics. Raw scores were used in preference to standard scores as the test had not been standardized in Hong Kong.

### Place value knowledge

Participants completed a number-comparison task identical to that used by Dowker et al. (2008), based on that of Donlan and Gourlay (1999). A pair of two-digit numbers was simultaneously presented to participants, who were asked to read them aloud and to point to the larger one within the pair. There were 24 pairs of numbers consisting of three types of number pairs: Transparent, Misleading, and Reversible. *Transparent* word pairs contained two numbers differing in the tens digit, thus requiring decade comparisons (e.g., 73 and 43) or contained repeated digits (e.g., 66 and 55). In *Misleading* number pairs, the smaller number contained a digit larger than the sum of digits in the larger item, (e.g., 51 and 47). *Reversible* pairs contained numbers whose tens and digit places were opposites (e.g., 85 and 58). An overall error score was calculated as in Dowker et al. (2008).

### Attitude toward mathematics (ATM)

Mathematics and Anxiety Questionnaire (MAQ; Thomas and Dowker, 2000) was used to measure children's ATM. Children answered four types of questions measuring self-perceived performance, attitudes in mathematics, unhappiness related to problems in mathematics, and anxiety related to problems in mathematics. There was a practice task followed by seven math-related situations: math in general, written calculations, mental calculations, easy calculations, difficult calculations, math homework, and listening and understanding the teacher during math lessons. Children answered on a 5-point scale using different pictures for each type of questions, such as ticks and crosses ("very good" to "very bad"), sweets and wasps ("like very much" to "hate very much"). The ratings varied from 0 for the most negative answer to 4 for the most positive answer, with a higher score indicating a more positive ATM. Overall the scale was found to be reliable (28 items,  $\alpha = 0.89$ ).

### Non-verbal intelligence

All children completed Raven's Colored Progressive Matrices Set A, AB, and B (Raven, 1962). Children were required to choose the correct answer from six options for 36 colored puzzles. Raven's tests are favored as a measure of nonverbal intelligence since they are considered "culture-fair," which is particularly important for cross-cultural studies. Raw scores were used in preference to standard scores as the available version of the test had not been standardized in Hong Kong.

## RESULTS

### DEMOGRAPHIC DATA

The means and SD of age, BAS total and arithmetic scores, Raven's matrices, MAQ, and Number Comparison total error scores of the different Schools (language groups) are shown in **Tables 1** and **2**. The variables were normally distributed, allowing subsequent parametric analyses. Participants with a Raven's score two SD away from the group mean were excluded.

### NONVERBAL INTELLIGENCE

Univariate ANOVA with School (three levels: HK-C, HK-E, UK) and Grade (two levels: first/second grade, third/fourth grade) as the independent variables (IV), and Raven's score as the dependent variable (DV) was conducted to investigate whether students differed in intellectual functioning. Children at the three schools differed significantly on Raven's matrices score,  $F(2,138) = 37.81$ ,  $p < 0.001$ ,  $\eta^2 = 0.36$ . *Post hoc* LSD tests revealed that the difference was driven by the UK school and the Hong Kong schools ( $p < 0.001$ ). The UK students had a lower score ( $\mu = 25.27$ ) than HK-E students ( $\mu = 29.98$ ) and HK-C students ( $\mu = 31.06$ ), while the Hong Kong schools did not differ significantly from each other. Children in the two grades were also significantly different from each other:  $F(1,138) = 7.43$ ,  $p = 0.007$ ,  $\eta^2 = 0.053$ . Third/fourth grade students performed better than first/second grade students.

There was no interaction of grade and school. Group differences in nonverbal intelligence were statistically controlled in subsequent analysis.

### COUNTING

For the Counting task, a successful attempt was one in which no mistakes were made. Hesitations of over three seconds, incorrect sequence or missing numbers constituted a failed attempt. Chi-squared contingency tests revealed a non-significant relationship between Schools and Success/Failure on the Counting Forward task for first-/second-graders. However, there was a significant relationship between Schools and Success/Failure on the task for third-/fourth-graders,  $X^2(2, N = 66) = 9.82$ ,  $p = 0.007$ . **Figure 1** depicted percentage of students who failed the Counting Forward task. Chi-squared test results for Counting Backward tasks revealed a significant relationship between Schools and Success/Failure on the task for first-/second-graders,  $X^2(2, N = 73) = 9.45$ ,  $p = 0.009$ . There was also a significant relationship between Schools and Success/Failure for third-/fourth-graders,  $X^2(2, N = 66) = 7.14$ ,  $p = 0.028$ . In both Counting Forward and Backward tasks, paired comparisons between groups were not possible due to relatively small sample sizes. **Figure 2** depicted percentage of students who failed the Counting Backward task.

### NUMBER COMPARISON

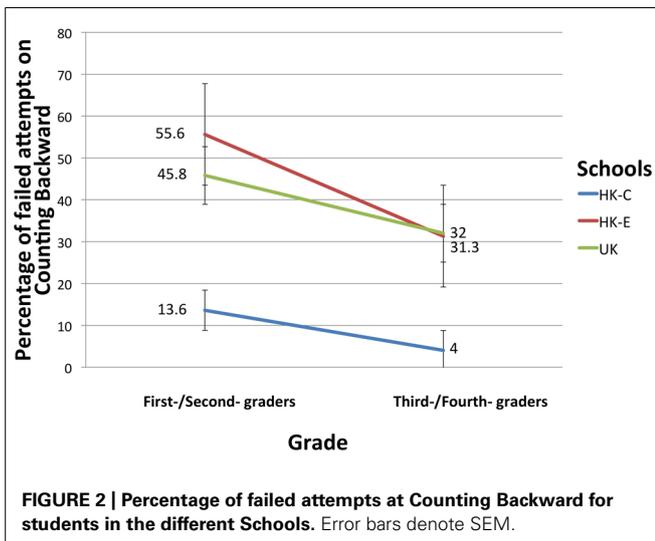
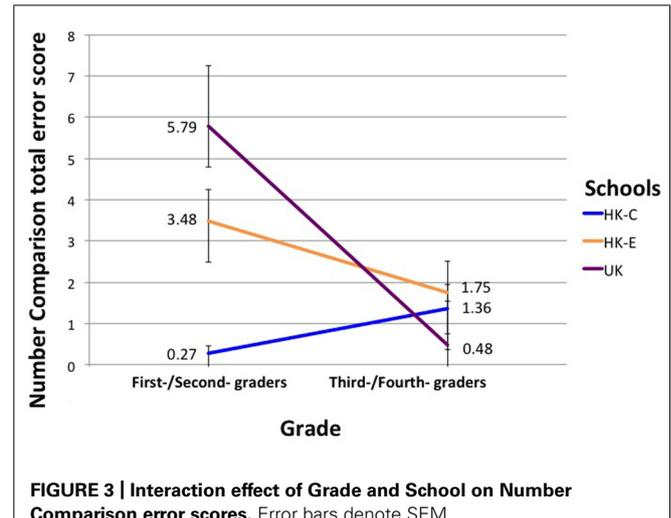
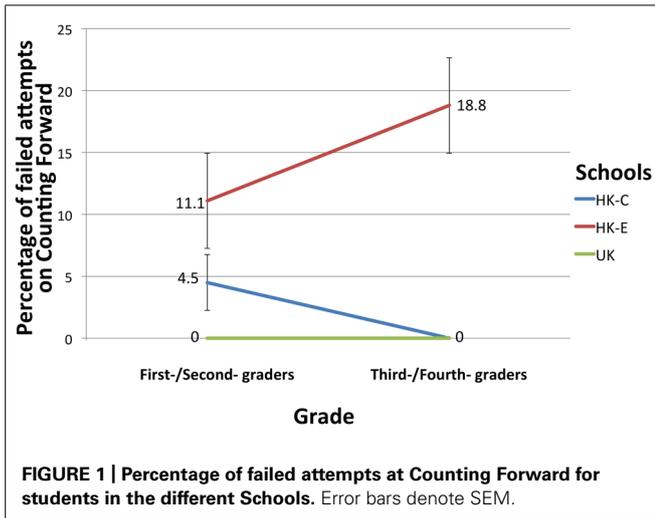
Univariate ANCOVA with Grade and School as IV, Number Comparison total error score as DV and Raven's matrices score as a covariate showed that Grade  $F(2,132) = 7.92$ ,  $p = 0.001$ ,  $\eta^2 = 0.11$ , as plotted in **Figure 3**. *Post hoc* pairwise comparisons showed that HK-C students were significantly better than HK-E and UK students in first/second grade:  $F(2,132) = 7.168$ ,  $p = 0.001$ ,  $\eta^2 = 0.098$ , but not in third/fourth grade. Also, within the UK group, first-/second-graders had higher error scores than third-/fourth-graders:  $F(1,132) = 19.007$ ,  $p < 0.001$ ,  $\eta^2 = 0.13$ .

**Table 1 | Mean age, BAS total score, and arithmetic scores, Raven's Matrices score, MAQ score, and Number Comparison task total error score for first/second grade students (SD in brackets).**

Schools	N	Age	BAS		Raven	MAQ	Number comparison
			Total	Arithmetic			
HK-C	22	6.95 (0.21)	18.09 (5.41)	11.95 (4.04)	29.95 (3.00)	49.68 (10.10)	0.27 (0.88)
HK-E	27	6.65 (0.48)	16.96 (4.10)	11.81 (3.97)	29.74 (2.77)	67.67 (24.99)	1.44 (1.58)
UK	24	6.85 (0.29)	9.21 (3.38)	4.33 (2.33)	24.17 (3.84)	71.04 (14.28)	3.08 (3.88)

**Table 2 | Mean age, BAS total score, and arithmetic scores, Raven's Matrices score, MAQ score, and Number Comparison task total error score for third/fourth grade students (SD in brackets).**

Schools	N	Age	BAS		Raven	MAQ	Number comparison
			Total	Arithmetic			
HK-C	25	8.96 (0.35)	26.92 (2.83)	22.24 (3.96)	32.04 (3.13)	46.84(3.67)	1.36 (9.82)
HK-E	16	8.56 (0.51)	23.62 (3.85)	18.69 (3.00)	30.38 (3.24)	58.56 (12.54)	1.75 (2.82)
UK	25	8.88 (0.31)	16.28 (2.70)	10.22 (2.58)	26.32 (4.44)	64.76 (13.61)	0.48 (0.96)

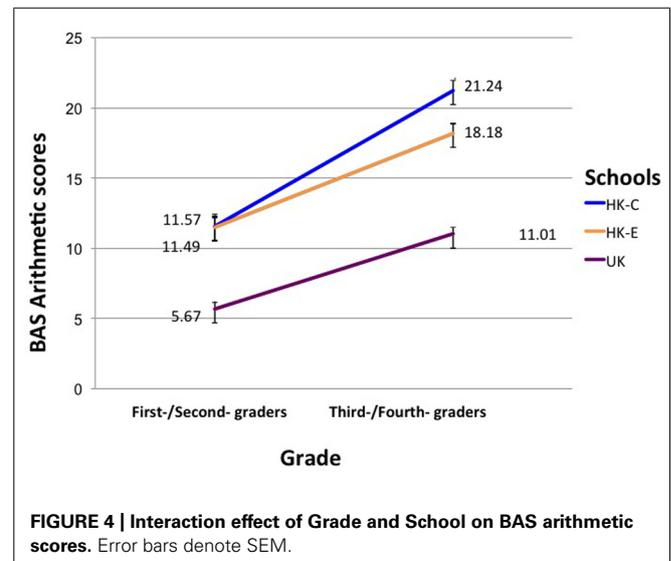


interaction between Grade and School, for either BAS total score or MAQ.

To investigate linguistic influences on arithmetic abilities specifically, a MANCOVA with Grade and School as IV, MAQ, and BAS arithmetics as DV and Raven’s matrices score as a covariate was conducted. Results showed significant main effects of Grade and BAS arithmetic score:  $F(1,130) = 154.34, p < 0.001, \eta^2 = 0.54$ , as well as School and BAS arithmetics:  $F(2,130) = 49.36, p < 0.001, \eta^2 = 0.43$ . A significant interaction was found between Grade and School for BAS arithmetics:  $F(2,130) = 5.31, p = 0.006, \eta^2 = 0.76$ . The interaction is plotted in **Figure 4**. First-/second-graders in the two Hong Kong schools did not significantly differ from each other in BAS arithmetic, but the UK first-/second-graders performed worse than the HK students. In third-/fourth-graders, all the schools differed in performance, with HK-C students performing better than HK-E students, who were in turn better than UK students in arithmetics.

**NUMERICAL SKILLS AND ATTITUDES**

To investigate the effect of linguistic influences on numerical skills, a MANCOVA with Grade and School as IV, MAQ, and BAS total scores as DV and Raven’s matrices score as a covariate showed a significant effect of Grade and BAS total scores:  $F(1,132) = 118.54, p < 0.001, \eta^2 = 0.47$ . Third/fourth grade students performed better on the BAS total scores than first/second grade students even after controlling for IQ. A significant effect of School and BAS total scores was found:  $F(2,132) = 41.98, p < 0.001, \eta^2 = 0.39$ . *Post hoc* tests revealed that BAS total scores from all schools significantly differed from each other, even after controlling for IQ. HK-C students performed the best, followed by HK-E students and then UK students. There was also a significant effect of Grade and MAQ:  $F(1,132) = 5.10, p = 0.026, \eta^2 = 0.04$ , with first/second grade students having a higher MAQ score. There was also a significant effect of School and MAQ:  $F(2,132) = 16.07, p < 0.001, \eta^2 = 0.20$ . *Post hoc* tests revealed that HK-C students scored lower than HK-E students and UK students, but HK-E students did not differ significantly from UK students. There was no significant



To investigate whether group differences in MAQ might be either exaggerating or masking group differences in arithmetical performance, a final ANCOVA was carried out with Grade and School as IV, BAS arithmetic as DV, and both Ravens and MAQ as covariates. There was again a significant effect of both Grade [ $F(1,129) = 33.44$ ;  $p < 0.026$ ;  $\eta^2 = 0.94$ ] and of School [ $F(2,129) = 12.42$ ;  $p < 0.059$ ;  $\eta^2 = 0.92$ ], with a significant interaction between School and Grade:  $F(2,129) = 5.01$ ;  $p = 0.008$ ,  $\eta^2 = 0.072$ . Ravens continued to be a significant covariate of Grade [ $F(1,129) = 12.85$ ;  $p < 0.001$ ;  $\eta^2 = 0.091$ ] but MAQ was not.

### PARENTAL INVOLVEMENT AND OUTSIDE SCHOOL MATHEMATICS

In order to establish whether parental involvement had an influence on the students' performance on BAS in the different schools, a chi-squared contingency test was conducted between Parental Involvement and Schools for first-/second-graders and third-/fourth-graders separately. No significant results were found in either grade, indicating that parental involvement in math did not differ significantly across the schools.

Similarly, to establish whether outside school mathematics instruction might have an influence on the students' performance in BAS in different schools, a chi-squared contingency test was carried out between Outside School Mathematics and Schools for first-/second-graders and third-/fourth-graders separately. No significant results were found in either grade. Therefore, the students' outside school formal mathematics instruction in math did not differ significantly across the three schools.

## DISCUSSION

### GENERAL DISCUSSION

This study aimed at examining the effect of exposure to the transparent Chinese counting system on counting, place value understanding general mathematical performance, and arithmetics through a cross-cultural study of HK-C, HK-E, and UK children. Our experimental design allows us to posit that if children performed in a descending order of HK-C > HK-E > UK, formal instruction and use of Mathematics in Chinese might be driving the difference in performance. If Hong Kong schoolchildren performed as a group performed better than UK children, however, it could be suggested that mere exposure or knowledge of Chinese counting system was enough to impact performance. The effect of duration of use of a Chinese counting system was also studied through the comparison of younger (first/second grade) and older (third/fourth grade) children. Our results showed that CNA imposed an effect on general mathematical abilities (BAS total scores). However, this effect was not apparent in relation to number representation (Counting Forward/Backward), or arithmetic abilities (BAS arithmetic scores) specifically. Furthermore, exposure to Chinese counting systems was only found to impact place value knowledge (Number Comparison) in younger, but not older, children. This suggested that while exposure to a regular counting system could advance place value understanding and general numerical abilities, it is unlikely to confer long-term benefits nor be able to explain cross-national differences in arithmetic abilities specifically.

In an attempt to exclude confounding variables to examine linguistic influences, our study minimizes the effects of known

alternative explanations to CNA in explaining number representations and arithmetic performance, such as IQ, ATM, outside school mathematics classes and parental involvement in children's learning.

### IQ

The UK students in our sample had a lower Raven's score than the Hong Kong groups. Since nonverbal IQ is related to general academic abilities including arithmetics, this effect was controlled as a covariate. Given that the correlations and group comparisons all controlled for the effect of IQ, significant differences obtained were explicable by some factors over and beyond the influence of IQ.

### Attitude toward mathematics

It has been suggested that more positive attitudes in Chinese compared to American students might contribute to the former putting more effort into their learning (Wong et al., 2001). However, our results showed that UK students indicated a more positive ATM, followed by HK-E and then HK-C students, which is the exact opposite of their pattern of performance on the arithmetic test. Our findings were in line with previous findings that students in countries ranking high in international comparisons disliked mathematics (Leung et al., 2006; Hirabayashi, 2006). Moreover, when attitude score was included as a covariate in the analysis of the effects of group and grade, it was found neither to affect BAS scores, nor to change the nature of the group differences. We cannot, however, exclude the possibility that Hong Kong students were more motivated to do well in academic assessments in general. It has previously been suggested that Chinese students were driven by pleasure derived from a result of the success attained in exams rather than through the process of learning *per se* (Leung et al., 2006).

### Outside school mathematics classes

Some students enrolled in tutorial classes outside of school. These tutorial classes are not usually subject-specific for primary school students, though some are (e.g., the Kumon Educational maths program). Although it could be expected that exposure and drilling in arithmetics might impact positively on arithmetic tests, extra-curricular mathematics class participation did not differ significantly across the three schools in our study. Hence, outside school mathematics exposure is unlikely to be responsible for the cross-cultural differences in arithmetic abilities found in our study, though one cannot rule out possible influences of more specific characteristics of the extracurricular instruction provided to different children.

### Parental involvement in children's education

Some studies suggested that Chinese parents were more involved in their children's education, giving more help or reprimand (Chen and Stevenson, 1989). Thus we asked students if their parents helped with their mathematics homework or taught them mathematics at home. However, our results showed that there was no statistically significant difference in self-reported level of parental involvement across the schools. Hence differences in arithmetic abilities found in this study were not likely to be due to disparate

parental involvement, though it is always necessary to be cautious about self-report measures.

### **Curriculum and educational system**

The primary educational system in Hong Kong is modeled on that of the British system, reflecting its colonial history to the UK. In Hong Kong, children receive primary education 'Primary 1 – Primary 6' from the ages of 6 until 12. In England, primary education spans over a similar age range, and is divided into 'Key Stage 1' (5–7 years old) and 'Key Stage 2' (7–11 years old). In both education systems, schools are required to teach a curriculum set by the government. The HK-C and HK-E students shared the same curriculum and Confucian traditions for academic excellence. In Hong Kong, more than half of primary school children were allocated centrally to Chinese-medium or English-medium schools. Hence, selection bias in relation to medium of instruction was unlikely to severely undermine our results. While the curriculum difference between Hong Kong and the UK was not possible to control, Tsang and Rowland (2005) had concluded that the two curricula were similar in content and organization. However, the scope of the study did not permit detailed comparisons of the implemented curricula and classroom teachings across the schools.

There are also other possible differences between the schools, which could have conceivably affected the results. Although there was no explicit difference in prestige or selectivity between the schools, and they were in similar neighborhoods, it is still possible that there might have been subtle differences between the parents, who chose to send their children to the Chinese- and English-medium schools. For example, the parents, who sent their children to the Chinese-medium school, might have identified more closely with Chinese culture, including an emphasis on mathematics and science. It may also be that some of the parents, who sent their children to the English-medium school, may have been responding to lower perceived mathematical ability in the children, by sending them to a school where they might compensate by acquiring fluency in a foreign language. The fact that the two Hong Kong groups did not differ in Ravens score reduces the likelihood that the differences in mathematical performance were due to some important pre-existing differences in ability; but one cannot rule out such differences altogether.

### **COUNTING**

Interestingly, UK students were found to be better than Hong Kong students at forward counting from 1 to 30 in first/second grade. This was inconsistent with the idea that regular number systems required less cognitive effort to learn and thus should be learnt earlier (Miller et al., 1995; Towse and Saxton, 1998). The observably poorer performance of HK-E students in forward counting across grades highlighted the caveat that it could not be determined whether the HK-E students could be counting or reading the numbers in Chinese in their heads and then giving an English response. Hence, their poor performance could be due to having to give response in a second language, especially one which is less transparent. It should be noted that it would be inevitably difficult to obtain a sample with no 'contamination' of

a second language in any likely setting for a bilingual educational system.

In Backward Counting from 30 to 1, Hong Kong students appeared to perform better than their HK-E and UK peers, but the difference between HK-E and UK children was minimal. Taking together the results of Forward and Backward Counting, it could be suggested that Forward Counting consisted of rote learning of the sounds of number strings; hence it might not be a real indication of children's number counting ability. When children were asked to count backward, which was much less common to hear and produce, the results showed a difference between children learning to count with a regular Chinese counting system and irregular English counting system. Since Chinese has a more transparent counting system, it is easy to infer the next number up or down the number line. Thus, students who learnt to count in Chinese could easily produce the backward sequence on the spot. In contrast, children who learnt to count in English had more difficulty, as it required 'flipping over' their phonological representation of the number strings.

### **NUMBER COMPARISON**

Our finding that HK-C students were significantly better than HK-E and UK students on the number comparison task in younger children but not older children suggested that transparency of the Chinese counting system might give children a 'head-start' in place value understanding. However, such an advantage bestowed by the CNA on first/second grade children was not 'sustainable,' as students who learn to count in irregular English counting system gradually 'caught up' in place value knowledge, as shown by the non-significant difference in the number comparison task across schools in third/fourth grade. That being said, such a conclusion is limited by a cross-sectional design and needs to be clarified in a longitudinal study in which children from HK-C, HK-E, and UK are followed through from first/second grade to third/fourth grade on the same task. It would also be interesting to replicate our study with an addition of a more explicit measure of place value knowledge (e.g., base-ten blocks) than our Number Comparison task. Our results support the practice of teaching young children from irregular counting systems to learn how numbers are formed in transparent number systems. Such an experience could serve both as cultural exposure and as a means to gain insight into the base system and place values.

### **MATHEMATICAL COMPETENCE**

Numerical competence was tested with the British Abilities Scale Number Skills Test, which was developed for students following the UK curriculum. Despite this potential advantage to the UK students, they performed the worst out of the three groups. Our results revealed an expected descending order of performance (HK-C, HK-E, UK) on general mathematic performance as measured by the total score on the BAS Number Skills test, as well as arithmetic performance in older children. Interestingly, however, such a disparity was not found for questions tapping arithmetic operations in younger children. HK-C did not achieve better arithmetic performance relative to HK-E children in the first/second grade. However, Hong Kong students as a whole still performed better than children in the UK.

As noted above, one caveat was that HK-E students could be disadvantaged by having to learn and respond in a second language. However, Dowker et al. (2008) showed that the advantages of learning Mathematics in Welsh held even if it was not a child's first or only language. Hence, the poorer performance of HK-E relative to HK-C children was unlikely to be due to disadvantages of learning in a second language. Although our results could be interpreted to mean that some exposure to a regular Chinese system was still advantageous even if it was not the formal medium of instruction at school, our results weaken the CNA *per se* as an explanation for better arithmetic abilities of Asian students.

## CONCLUSION AND FUTURE STUDIES

In conclusion, this study demonstrated that young children who were learning mathematics in Chinese were better at manipulating the number line than those learning mathematics in English, whether English be their first or second language. We also showed that linguistic transparency in number representations might facilitate place value learning in young children, but such an advantage is neither sustainable nor necessarily translated to better arithmetic performance in older children. Our pattern of findings replicated that of Dowker et al. (2008), whereby children who learnt mathematics in regular counting system out-performed those who learnt mathematics in English on the Number Comparison task but not (at least for the younger children) on a test of more general arithmetic. The mechanism underlying the linguistic influence is, however, yet to be elucidated. As yet, the evidence is not sufficient to demonstrate that the CNA, as framed in terms of transparency of numbers, can explain the cross-national differences in arithmetic consistently demonstrated across age groups. The fact that children in HK-E performed significantly better than the UK children, although both groups were educated in English, suggests that general educational and cultural differences are at least as important as linguistic differences; though one cannot rule out the possibility that the HK-E children were advantaged by their exposure to Chinese counting at home.

More research is needed to fully understand the nature and extent of the differences in arithmetic between Chinese- and English-speaking children. To date, only a few studies have taken advantage of the unique opportunities afforded by the Chinese- and English-medium of instruction to tap linguistic influence in mathematics learning. In an ideal world, children of similar backgrounds would be randomly assigned to Chinese versus English medium schools, to rule out any effects of self-selection. In practice, this would of course be impossible. However, extending the number of schools studied would reduce the chances of the results being due to sample or school characteristics that are unrelated to language. It would also be interesting if the study could be extended to even younger children in kindergarten in order to test for even earlier effects. Moreover, it would be desirable to include a wider variety of number representation tasks: for example, including the blocks task of Miura et al. (1988). It is a potential limitation that the British Abilities Scales and the Raven's Matrices were developed for use in Britain, rather than in Hong Kong. The fact that Hong Kong pupils outperformed British pupils on both tests makes it in fact unlikely that these tests

involved unfamiliar or unsuitable material for use in Hong Kong schools. However, future studies should also attempt to develop and standardize tests for simultaneous use in the UK and in Hong Kong.

## REFERENCES

- Ackerman, P. L. (1988). Determinants of individual differences during skill acquisition: cognitive abilities and information processing. *J. Exp. Psychol. Gen.* 117, 288–318. doi: 10.1037/0096-3445.117.3.288
- Aunio, P., Aubrey, C., Godfrey, R., Pan, Y., and Liu, Y. (2008). Children's early numeracy in England, Finland and People's Republic of China. *Int. J. Early Years Educ.* 16, 203–221.
- Aunio, P., Ee, J., Lim, S. E. A., Hautamaki, J., and Van Luit, J. E. H. (2004). Young children's number sense in Finland, Hong Kong and Singapore. *J. Early Years Educ.* 12, 195–216. doi: 10.1080/0966976042000268681
- Aunio, P., Niemivirta, M., Hautamaki, J., Van Luit, J. E. H., Si, J., and Zhang, M. (2006). Young children's number sense in China and Finland, *Scand. J. Educ. Res.* 50, 482–502. doi: 10.1080/00313830600953576
- Borghi, A. M., and Binkofski, F. (2014). *Words as Social Tools: An Embodied View on Abstract Concepts*. Berlin, NY: Springer. doi: 10.1007/978-1-4614-9539-0
- Borghi, A. M., Flumini, A., Cimatti, F., Marocco, D., and Scorolli, C. (2011). Manipulating objects and telling words: a study on concrete and abstract words acquisition. *Front. Psychol.* 2:15. doi: 10.3389/fpsyg.2011.00015
- Cankaya, O., LeFevre, J. A., and Dunbar, K. (2014). The role of number naming systems and numeracy experiences in children's rote counting: evidence from Turkish and Canadian children. *Learn. Individ. Differ.* 32, 238–245. doi: 10.1016/j.lindif.2014.03.016
- Chen, C., and Stevenson, H. W. (1989). Homework: a cross-cultural examination. *Child Dev.* 60, 551–561. doi: 10.2307/1130721
- Cheng, Z. J., and Chan, L. K. S. (2005). Chinese number-naming advantages? Analyses of Chinese pre-schoolers' computational strategies and errors. *Int. J. Early Years Educ.* 13, 179–192. doi: 10.1080/09669760500171279
- Dehaene, S. (1997). *The Number Sense: How the Mind Creates Mathematics*. Great Britain: The Penguin Press.
- Donlan, C., and Goulay, S. (1999). The importance of nonverbal skills in the acquisition of place-value knowledge: evidence from normally-developing and language-impaired children. *Br. J. Dev. Psychol.* 17, 1–19. doi: 10.1348/026151099165113
- Dowker, A., Bala, S., and Lloyd, D. (2008). Linguistic influences on mathematical development: how important is the transparency of the counting system. *Philos. Psychol.* 21, 523–538. doi: 10.1080/09515080802285511
- Dowker, A., and Lloyd, D. (2005). "Linguistic influences on numeracy," in *Mathematics in the Primary School*, eds D. V. Jones, A. Dowker, and D. Lloyd (Oxford: Bangor School of Education), 21–35.
- Edgeworth, M., and Edgeworth, R. L. (1798). *Practical Education*. London: J. Johnson.
- Fuson, K. C., and Kwon, Y. (1991). "Chinese-based regular and European irregular systems of number words: the disadvantages for English-speaking children," in *Language and Mathematical Education*, eds K. Durkin and B. Shire (Milton Keynes: Open University Press), 211–226.
- Fuson, K. C., and Kwon, Y. (1992). "Learning addition and subtraction: effects of number word and other cultural tools," in *Pathways to Number: Children's Developing Numerical Abilities*, eds J. Bideaud, C. Meljac, and J. Fischer (Hillsdale, NJ: Erlbaum), 283–306.
- Gentner, D., and Boroditsky, L. (2001). "Individuation, relativity and early word learning," in *Language. (Acquisition) and Conceptual Development*, eds M. Bowerman and S. Levinson (New York: Cambridge University Press), 215–256. doi: 10.1017/CBO9780511620669.010
- Göbel, S. M., Shaki, S., and Fischer, M. H. (2011). The cultural number line: a review of cultural and linguistic influences on the development of number processing. *J. Cross Cult. Psychol.* 42, 543–565. doi: 10.1177/0022022111406251
- Hirabayashi, I. (2006). "A traditional aspect of mathematics education in Japan," in *Mathematics education in different cultural traditions: a comparative study of East Asia and the West. The 13th ICMI Study*, eds F. K. S. Leung, K. D. Graf, and F. J. Lopez-Real (Dordrecht: Springer.), 51–64.

- Klein, E., Bahnmüller, J., Mann, A., Pixner, S., Kaufmann, L., Nuerk, H. C., et al. (2013). Language influences on numerical development—Inversion effects on multi-digit number processing. *Front. Psychol.* 4:480. doi: 10.3389/fpsyg.2013.00480
- Krinzinger, H., Gregoire, J., Desoete, A., Kaufmann, L., Nuerk, H., and Willmes, K. (2011). Differential language effects on numerical skills in second grade. *J. Cross Cult. Psychol.* 42, 614–629. doi: 10.1177/0022022111406252
- Leung, F. K. S., Graf, K. D., and Lopez-Real, F. J. (eds). (2006). *Mathematics Education in Different Cultural Traditions: a Comparative Study of East Asia and the West. The 13th ICMI Study*. Dordrecht: Springer. doi: 10.1007/0-387-29723-5
- MacLean, M., and Whitburn, J. (1996). “Number name systems and children’s early number knowledge: a comparison of Welsh and English speaking children,” in *Proceedings of the XVIIth Biennial International Society for the Study of Behavioral Development (ISSBD) Conference*, Quebec City, QC, Canada.
- Miller, K. F., Kelly, M., and Zhou, X. (2005). “Learning mathematics in China and the United States: Cross-cultural insights into the nature and course of preschool mathematical development,” in *Handbook of Mathematical Cognition*, ed. J. I. D. Campbell (New York: Psychology Press), 163–178.
- Miller, K. F., Smith, C. M., Zhu, J., and Zhang, H. (1995). Preschool origins of cross-national differences in mathematical competence: the role of number-naming systems. *Psychol. Sci.* 6, 56–60. doi: 10.1111/j.1467-9280.1995.tb00305.x
- Miura, I. T., Kim, C. C., Chang, C. M., and Okamoto, Y. (1988). Effects of language characteristics on children’s cognitive representation of numbers: cross-national comparisons. *Child Dev.* 59, 1445–1450. doi: 10.2307/1130659
- Miura, I. T., and Okamoto, Y. (2003). “Language supports for mathematics understanding and performance,” in *The Development of Arithmetic Concepts and Skills: Constructing adaptive expertise. Studies in mathematics thinking and learning*, eds A. J. Baroody and A. Dowker (Mahwah, NJ: Erlbaum), 229–242.
- Miura, I. T., Okamoto, Y., Kim, C. C., Steere, M., and Fayol, M. (1993). First graders’ cognitive representation of number and understanding of place value: cross-national comparisons – France, Japan, Korea, Sweden, and the United States. *J. Educ. Psychol.* 85, 24–30. doi: 10.1037/0022-0663.85.1.24
- Mullis, I., Martin, M., and Foy, P. (2008). *TIMSS 2007 International Mathematics Report: Findings from IEA’s International Mathematics and Science Study at the Fourth and Eighth Grades*. Boston, MA: Centre for the Study of Testing, Evaluation and Educational Policy, Boston College.
- Mullis, I., Martin, M., Gonzalez, E., Garden, R., O’Connor, K., Chrostowski, S., et al. (2000). *TIMSS 1999: International Mathematics Reports, Finding from IEA’s Repeat of Third International Mathematics and Science Study at the Eighth Grade*. Boston, MA: Center for the Study of Testing, Evaluation and Educational Policy, Boston College.
- Ng, S. S. N., and Rao, N. (2010). Chinese Number words, culture, and mathematics learning. *Rev. Educ. Res.* 80, 180–206. doi: 10.3102/0034654310364764
- Pixner, S., Kaufmann, L., Moeller, K., Hermanova, V., and Nuerk, H.-C. (2011). Whorf reloaded: language effects on non-verbal number processing in 1st grade – a trilingual study. *J. Exp. Child Psychol.* 108, 371–382. doi: 10.1016/j.jecp.2010.09.002
- Provasnik, S., Kastberg, D., Ferraro, D., Lemanski, N., Roey, S., and Jenkins, F. (2012). *Highlights from TIMSS 2011: Mathematics and Science Achievement of U.S. Fourth- and Eighth-Grade Students in an International Context*. Washington, DC: National Center for Education Statistics.
- Rasmussen, C., Ho, E., Nicoladis, E., Leung, J., and Bisanz, J. (2006). Is the Chinese number-naming system transparent? Evidence from Chinese-English bilingual children. *Can. J. Exp. Psychol.* 60, 60–67.
- Raven, J. C. (1962). *Coloured Progressive Matrices Sets A, AB, B*. London: H.K. Lewis.
- Saxton, M., and Towse, J. N. (1998). Linguistic relativity: the case of place value in multi-digit numbers. *J. Exp. Child Psychol.* 69, 66–79. doi: 10.1006/jecp.1998.2437
- Stedman, L. C. (1997). International achievement differences: an assessment of a new perspective. *Educ. Res.* 26, 4–15.
- Stevenson, H. W., Chen, C., and Lee, S. Y. (1993). Mathematics achievement of Chinese, Japanese and American children: ten years later. *Science* 259, 53–58. doi: 10.1126/science.8418494
- Thomas, G., and Dowker, A. (2000). “Mathematics anxiety and related factors in young children,” in *Proceedings of the Developmental Section Conference* (Bristol: British Psychological Society).
- Towse, J. N., and Saxton, M. (1998). Linguistic influences on children’s number concepts: methodological and theoretical considerations. *J. Exp. Child Psychol.* 66, 362–375. doi: 10.1006/jecp.1997.2389
- Tsang, F. W. K., and Rowland, T. (2005). The subject matter knowledge of Hong Kong Primary School mathematics teachers. *Paper Presented at the European Conference on Educational Research*, Univeristy College Dublin, Dublin, OH, 7–10.
- Wang, J., Lin, E., Tanase, M., and Sas, M. (2008). Revisiting the influence of numerical language characteristics on mathematics achievement: comparison among China, Romania, and US. *IEJME* 3, 1–23.
- Wong, N. Y., Lam, C. C., Wong, K. M. P., Leung, F. K. S., and Mok, I. A. C. (2001). Students’ views of mathematics learning: a cross-sectional survey in Hong Kong. *Educ. J.* 29, 38–59.
- Zhao, D., and Singh, M. (2011). Why do Chinese-Australian students outperform their Australian peers in mathematics: a comparative case study. *IJSME* 9, 87.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 20 November 2014; accepted: 09 February 2015; published online: 26 February 2015.

Citation: Mark W and Dowker A (2015) Linguistic influence on mathematical development is specific rather than pervasive: revisiting the Chinese Number Advantage in Chinese and English children. *Front. Psychol.* 6:203. doi: 10.3389/fpsyg.2015.00203  
This article was submitted to *Developmental Psychology*, a section of the journal *Frontiers in Psychology*.

Copyright © 2015 Mark and Dowker. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Corrigendum: Linguistic influence on mathematical development is specific rather than pervasive: revisiting the Chinese Number Advantage in Chinese and English children

Winifred Mark<sup>1</sup> and Ann Dowker<sup>2\*</sup>

<sup>1</sup> Department of Psychology, University of Hong Kong, Hong Kong, China, <sup>2</sup> Department of Experimental Psychology, Oxford University, Oxford, UK

**Keywords:** linguistic transparency, counting system, arithmetic, cross-cultural, Chinese Number Advantage

## A corrigendum on

### Linguistic influence on mathematical development is specific rather than pervasive: revisiting the Chinese Number Advantage in Chinese and English children

by Mark, W., and Dowker, A. (2015). *Front. Psychol.* 6:203. doi: 10.3389/fpsyg.2015.00203

Due to an oversight, the two sentences preceding the final sentence in the abstract should be changed to read: Results indicated that students in HK-C were better at counting backward than those in HKE, who were in turn better than the UK students. However, there was no statistical difference in counting forward or place value understanding. Children in both Hong Kong schools performed better at the arithmetic test than the UK children. Among the older group, the HK-C children performed better on the arithmetic test than the HK-E children, but no such difference was found in the younger group.

The authors apologize for this mistake.

This error does not change the scientific conclusions of the article in any way.

## AUTHOR CONTRIBUTIONS

All authors listed, have made substantial, direct and intellectual contribution to the work, and approved it for publication.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Mark and Dowker. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

## OPEN ACCESS

### Edited and reviewed by:

Yvette Renee Harris,  
Miami University, USA

### \*Correspondence:

Ann Dowker  
ann.dowker@psy.ox.ac.uk

### Specialty section:

This article was submitted to  
Developmental Psychology,  
a section of the journal  
Frontiers in Psychology

**Received:** 20 February 2016

**Accepted:** 24 February 2016

**Published:** 16 March 2016

### Citation:

Mark W and Dowker A (2016)

Corrigendum: Linguistic influence on  
mathematical development is specific  
rather than pervasive: revisiting the

Chinese Number Advantage in  
Chinese and English children.

*Front. Psychol.* 7:342.

doi: 10.3389/fpsyg.2016.00342



# How space-number associations may be created in preliterate children: six distinct mechanisms

Hans-Christoph Nuerk<sup>1,2,3\*</sup>, Katarzyna Patro<sup>2,4</sup>, Ulrike Cress<sup>1,2,3</sup>, Ulrike Schild<sup>1</sup>, Claudia K. Friedrich<sup>1</sup> and Silke M. Göbel<sup>5</sup>

<sup>1</sup> Department of Psychology, Eberhard Karls University of Tuebingen, Tuebingen, Germany

<sup>2</sup> Leibniz Institute for Knowledge Media, Knowledge Media Research Center, Tuebingen, Germany

<sup>3</sup> LEAD Graduate School, Eberhard Karls University of Tuebingen, Tuebingen, Germany

<sup>4</sup> Department of Psychology, University of Warsaw, Warsaw, Poland

<sup>5</sup> Department of Psychology, University of York, York, UK

## Edited by:

Emily Mather, University of Hull, UK

## Reviewed by:

Ruth Ford, Anglia Ruskin University, UK

Jeffrey Coldren, Youngstown State University, USA

Samuel Shaki, Ariel University, Israel

## \*Correspondence:

Hans-Christoph Nuerk, Department of Psychology, Eberhard Karls University of Tuebingen, Schleichstrasse 4, 72076 Tuebingen, Germany  
e-mail: hc.nuerk@uni-tuebingen.de

The directionality of space-number association (SNA) is shaped by cultural experiences. It usually follows the culturally dominant reading direction. Smaller numbers are generally associated with the starting side for reading (left side in Western cultures), while larger numbers are associated with the right endpoint side. However, SNAs consistent with cultural reading directions are present before children can actually read and write. Therefore, these SNAs cannot only be shaped by the direction of children's own reading/writing behavior. We propose six distinct processes – one biological and five cultural/educational – underlying directional SNAs before formal reading acquisition: (i) *Brain lateralization*, (ii) *Monitoring adult reading behavior*, (iii) *Pretend reading and writing, and rudimentary reading and writing skills*, (iv) *Dominant attentional directional preferences in a society, not directly related to reading direction*, (v) *Direct spatial-numerical learning*, (vi) *Other spatial-directional processes independent of reading direction*. In this mini-review, we will differentiate between these processes, elaborate when in development they might emerge, discuss how they may create the SNAs observed in preliterate children and propose how they can be studied in the future.

**Keywords:** space-number associations, reading acquisition, numerical development, literacy, preliteracy, SNARC, number acquisition

## THE READING AND WRITING DIRECTION ACCOUNT IN ADULTS

One of the most intriguing findings in the field of Numerical Cognition is that numbers in adults are automatically associated with a spatial horizontal dimension (Fischer and Shaki, 2014). In Western countries, relatively larger numbers are usually associated with the right side in space and smaller numbers with the left side in space. The most widely studied demonstration of such an association is the so-called SNARC-effect (Spatial-Numerical Association of Response Codes; Dehaene et al., 1993): even in tasks in which number magnitude is irrelevant (e.g., parity judgment tasks), participants are faster to respond to larger numbers with the right hand, and to smaller numbers with the left hand (Wood et al., 2008).

The *common reading account* proposes that the origin of this directionality stems from reading habits. Suggested by Dehaene et al. (1993), this account was further corroborated in a series of studies by [e.g., Shaki and Fischer (2008), Fischer et al. (2009), Shaki et al. (2009); see also Zebian (2005)]. They showed that general and situational exposure to right-to-left writing modulated or even reversed the common SNARC effect – participants exposed to right-to-left reading habits had a null or right-to-left SNARC effect. However, there are other accounts of the origin of SNAs. For instance, some researchers propose that the SNARC effect is created by the order of numbers in verbal working memory

sequences (e.g., van Dijck and Fias, 2011). Others suggest that the direction of the SNARC effect might be triggered by early finger counting habits (an embodied account; Fischer, 2008) or that verbal-linguistic markedness might contribute to number-parity and number magnitude representations (Nuerk et al., 2004). A detailed discussion of these accounts is beyond the scope of the current review; here, we will focus on the dominant account, which is the *common reading account*.

## SPACE-NUMBER ASSOCIATIONS IN CHILDREN

Space-number associations (SNAs) develop in early childhood (McCrink and Opfer, 2014). Western preschoolers have a strong preference for left-to-right object counting (Briars and Siegler, 1984; Opfer et al., 2010; Shaki et al., 2012; Knudsen et al., in press) as well as for left-to-right sequences of Arabic digits (Opfer and Furlong, 2011). In a typical counting task, an explicit spatial-numerical decision has to be made, i.e., to start from the left or from the right. However, preschoolers show SNAs even in tasks not requiring an explicit spatial-numerical decision. Patro and Haman (2012) observed a SNARC-like effect in a non-symbolic numerosity comparison task in children as young as 3- and 4-years-old. All these children were clearly preliterate, so their reading habits could not explain their SNAs. In addition, SNAs in preschool children are already automatic and present even when magnitude is not task-relevant. Hoffmann et al. (2013; Experiment 2) observed

a classical SNARC effect in children as young as 5;5 years when children had to decide whether Arabic numbers changed to red or to green, by pressing a left- or right-located button. A SNARC-like interaction between number magnitude and response side was observed. Thus, number magnitude was task-irrelevant (children had to decide about color), but automatically activated. Moreover, there was no explicit instruction that magnitude should be related to one side of space. The presence of SNAs in preschool children clearly challenges the common reading account for SNAs, because those children have not yet developed reading habits themselves.

Recently, de Hevia et al. (2014) observed that already 7-months-old infants, growing up in Italy<sup>1</sup> (left-to-right-reading), showed a preference for left-to-right increasing sequences of sets' numerosities. They proposed an alternative to the common reading account and suggested *biological predisposition* to cause SNAs in very young children. These biologically determined SNAs might later be modulated or even reversed by reading/writing acquisition.

Even such a *combination account* of biological left-to-right predisposition and later modulation by cultural reading habits is at odds with recent studies. Shaki et al. (2012) showed that reading/writing habits in a society modulated counting habits already in preliterate children. British 3–6-years-old preschool children counted mainly from left-to-right, whereas the majority of the Israeli and Palestinian children (growing up in right-to-left reading cultures) counted from right-to-left. The combination account cannot explain these data. Its biological component cannot explain any cultural variation by reading habits at all. Its reading experience component cannot explain cultural modulation before reading acquisition.

Spatial-directional training also shapes or modulates SNAs in preliterates. Patro et al. (in press) provided directional attentional non-numerical training to 3–4-years-old children. They observed that left-to-right training led to a subsequent left-to-right SNARC-like effect, while right-to-left training led to a right-to-left SNARC-like effect. In another study, Göbel et al. (2014) tested counting direction in British and Arab preschoolers before and after a 5 min reading-related experience that was either left-to-right or right-to-left. They found that, irrespective of children's initial counting direction, most children who observed left-to-right reading counted left-to-right, and most children who observed right-to-left reading counted right-to-left. Such modulation of SNA direction by training also speaks against an exclusively biological account.

Both studies clearly show that spatial-directional experience shapes SNAs in preschoolers. In addition, taking both studies together they make an important point, which will drive our review: different SNA types were modulated by different spatial (training) mechanisms. Patro et al. (in press) conducted an implicit attentional training, not related to reading observation, and this training affected an implicit directional measure of SNA (the SNARC effect), but did not lead to a change in explicit counting direction. Similarly, Göbel et al. (2014) showed an effect on explicit counting direction only when the training included explicit reading observation but not with implicit attentional

training. This is in line with Kamawar et al.'s (2010) observation that children have a strong idea which explicit SNA is correct. They showed that the majority of 5–11-years-old children they tested in Canada believed that the order in which items were counted was important. Most children favored a left-to-right, top-to-bottom order of counting. Thus, children are very aware of explicit counting direction and have a clear idea of what the 'correct' direction of counting is. For children, this 'correct' direction seems to be consistent with their particular cultural reading/writing habits.

There is now clear empirical evidence that SNAs can be formed in preschool children, but we still lack a coherent theoretical proposal that could explain which concrete mechanisms or processes contribute to the emergence of number-space effects in young children. This is an obvious gap in this line of research. This mini-review aims to close this gap by proposing and discussing *six distinct mechanisms*.

It is important to note that numbers can be linked to spatial directions in different ways. Patro et al. (2014), who proposed four SNAs in general, described two *spatial-directional* SNA types in particular:

- (i) Associations between cardinalities and spatial directions: in this SNA, there is a directional association similar like in a SNARC effect – in left-to-right reading cultures larger numerosities are responded to faster on the right side and smaller numerosities on the left.
- (ii) Associations between ordinalities and spatial directions: in this SNA, spatial direction is related to ordinality (e.g., the direction of counting) – it is not necessarily related to cardinality because younger preschoolers do not know that the end point of the counting sequence equals the cardinality (i.e., the total number of objects in the sequence).

The mechanisms outlined in this review may not contribute equally to the emergence of the two SNA types described above. These mechanisms, their differential impact, and the probable age of onset will be defined and systematically demarcated in the remainder of this review.

## MECHANISMS POTENTIALLY INDUCING SPATIAL-NUMERICAL DIRECTIONALITY IN PRELITERATE CHILDREN

### BRAIN LATERALIZATION

Brain lateralization may play an important role for early spatial-directional preferences (Rugani et al., in press, 2015, for animal studies). Directional spatial-numerical biases in 7-months-old infants have been interpreted as an innate disposition to associate larger numerosities with one side in space (de Hevia et al., 2014). While such findings may be explained by innate biases, they are not fully conclusive yet: first, so far, no evidence has been obtained that the spatial-numerical biases vary systematically with an indirect measure of brain lateralization: handedness. Second, early presence of a mechanism does not necessarily imply innateness – 7 months might be long enough to learn about spatial-directional regularities in a social cultural setting. Third, even spatial biases which seem to be strongly predisposed might be subject to cultural influences (Güntürkün, 2003; Shaki, 2013).

<sup>1</sup>Personal communication with Maria Dolores de Hevia.

To be clear, these arguments do not preclude a role of brain lateralization in humans but, in our opinion, the case is far from closed.

### MONITORING ADULT READING BEHAVIOR

Joint book reading activity promotes emergent literacy (including print awareness) in children who are not yet conventional readers (Sénéchal et al., 1996; Mol et al., 2009). Via joint book reading, preliterate children could learn about text directionality by observing their parents pointing to particular places in text or referring to subsequent pictures (Dobel et al., 2007; McCrink et al., 2011). Knowledge of spatial organization of script and pictures in books (and also about the organization of books) could be acquired very early because adults start reading books to children as young as 1–2 years (Sénéchal et al., 1995; Fletcher and Reese, 2005). So, by reading books to children, adults may impose an attentional directionality, which children internalize even before they formally acquire reading skills.

### PRETEND READING AND WRITING, AND RUDIMENTARY READING AND WRITING SKILLS

Children acquire basic aspects of reading and writing well before formal instruction in school starts (Snow et al., 1998). In pretend reading, typically developing children at the end of their third year not only demonstrate that they know how to hold a book and turn pages in their native writing system, but also that they know that stories progress as pages are turned and that a story has a beginning, middle and end (e.g., Doake, 1985; Sulzby, 1985; Valencia and Sulzby, 1991). Also, starting at the end of age 3, approximate word-by-word pointing in pretend reading can be observed (Dooley, 2010). In pretend writing, preliterate children ‘write’ lists, thank-you notes, etc. (Dyson, 1982). Thus, young children at least start extracting the characteristic direction of their native language’s writing system. Between the ages of 3 and 4 children become more and more aware of the elements of writing and their linearity so that most 4 years-old can read and write one or more simple words, including their own name (Hildreth, 1936; Bloodgood, 1999; Puranik et al., 2011, 2013). That is, the directional process related to the local writing system appears to become active at the end of the third year and further elaborated in older preschoolers.

### DOMINANT ATTENTIONAL-DIRECTIONAL PREFERENCES IN A SOCIETY, NOT DIRECTLY RELATED TO READING DIRECTION

Reading and writing habits may influence directional preferences which at first sight have nothing to do with reading and writing themselves. First, visuo-spatial processing appears to be biased by writing direction. For instance, Arabic participants preferred drawing horizontal lines from right-to-left, while English-speaking participants preferred drawing them from left-to-right (Lieblich et al., 1975). Culture-dependent line bisection biases have been observed both in adults (Chokron and Imbert, 1993; Kazandjian et al., 2010; Rinaldi et al., 2014) and preliterate preschoolers (Chokron and De Agostini, 1995; but see Fagard and Dahmen, 2003). Second, spatial imagery also appears to be biased by writing direction. Hindi participants, reading from left to right, drew bicycles or elephants facing to the left, whereas Arab

participants exhibited a rightward bias for those objects (Vaid, 1995). For temporal preferences (e.g., meals of the day), adults tended to prefer horizontal alignment corresponding to their reading habits, i.e., future to the right in left-to-right writing systems and future to the left in right-to-left writing systems (Tversky et al., 1991). Furthermore, spatial representations of actions appeared to be modulated by reading direction. Adults exposed to left-to-right writing systems preferentially place and expect agents on the left side of a picture, whereas adults exposed to right-to-left writing systems show the reverse pattern (Maass and Russo, 2003; Dobel et al., 2007; Maass et al., 2009). In sum, adults engage in all kinds of attentional-directional behaviors which are not directly related to reading/writing, but which are nevertheless consistent with the direction of reading/writing in a society. Children may observe such behaviors from parents and other models and imitate them.

Importantly, some culture-dependent spatial directional *actions* themselves do not develop before school: children of school age, but not preschoolers showed culture-dependent directionality in drawing (Kebbe and Vinter, 2013). Similarly, children of school age showed temporal ordering of spatial relations (Tversky et al., 1991), but preschoolers did not show a preference regarding spatial placement of agents (Chokron and De Agostini, 2000; Spalek and Hammad, 2005; Dobel et al., 2007; McCrink et al., 2014; for reviews see Kazandjian and Chokron, 2008; Chokron et al., 2009).

It should be also noted that many applications for electronic devices (computers, tablets, smartphones) are adjusted for different reading/writing directions. Even operating systems (e.g., Windows) have a Hebrew/Arabic version, which starts from right-to-left: the ‘start’ button is located on the right side of the screen and the window menu opens from right-to-left. Similar directional differences can be found in childrens’ applications /games, which are designed for 3–4-years-old kids, who are not yet able to read. Thus, via such applications, young children are directly exposed to certain attentional-directional cultural preferences<sup>2</sup>.

In sum, there are multiple cultural spatial-directional biases in everyday actions which are not directly related to reading behavior, but are nevertheless consistent with its directionality in the local culture. It is conceivable that such biases influence attentional directionality in preliterate children.

### DIRECT SPATIAL-NUMERICAL LEARNING

The mechanisms described above are concerned with spatial-directional biases which are not related to numbers. However, there are also direct explicit instructions of spatial-numerical relations. For example, children are exposed to certain spatial arrangements of numbers in their picture books, and they are often formally and informally taught to count objects in a certain order. Lindemann et al. (2011) have shown that finger-counting habits also seem to differ between cultures. Finger counting habits even strongly differ between cultures which have the same script [see Bender and Beller, 2012, for between culture-variations; Wasner et al. (in press), for within-culture variations]. Thus, there is a spatial-numerical component in finger counting that goes beyond reading directionality and which is directly learnt in a given culture.

<sup>2</sup>We thank a reviewer for pointing this out to us.

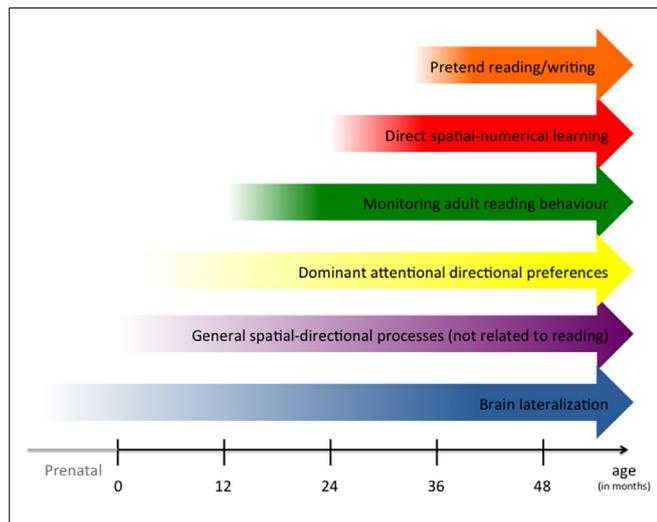
Therefore, children may directly learn certain directionalities of space-number relations from adult models or instruction. This direct instruction of SNAs may begin at about 2–3 years, when children start to count.

**OTHER SPATIAL-DIRECTIONAL PROCESSES INDEPENDENT OF READING DIRECTION**

Cultures may also differ in other spatial-directional processes, which are not related to reading direction or explicit numerical instruction. For instance, spatial looking behavior when crossing a street is influenced by the lane on which the traffic usually drives (first look to the right for left-lane traffic in the UK, first look to the left, for right-lane traffic in the rest of Europe). Such spatial-directional mechanisms might affect SNAs as well. However, we are not aware of any studies yet examining such influences. We would hypothesize that other spatial-directional influences generally increase directional SNAs when they are congruent to the cultural reading/writing direction and decrease SNAs when they are incongruent.

**WHERE WE ARE AND WHAT WE CAN CONCLUDE**

We have defined and delineated six distinct mechanisms which might be responsible for the emergence of spatial-numerical directional preferences before formal literacy (for an overview including time of onset, see **Figure 1**). These mechanisms are probably often consistent, but can be sometimes in conflict. For instance, an Arab parent may read Arab children’s books from right-to-left, but may count objects from left-to-right, because this is how numbers are ordered in most numerical and arithmetic graphs. Therefore, different SNA types may be represented in a different fashion or even in an opposite direction because they are learnt by different, possibly directionally conflicting, mechanisms.



**FIGURE 1 | Overview of the different mechanisms underlying the acquisition of spatial-numerical associations.** Mechanisms are ordered according to their probable age of onset according to the literature. Exact time of onset is often difficult to determine, therefore, the shaded start of the arrows depicts the probable range of onset in typically developing children. Note that brain lateralization starts before birth and that all mechanisms continue to activate spatial-numerical associations beyond the age of 48 months as indicated by the arrows.

Most of the learning mechanisms proposed here are related to embodied spatial-numerical learning (e.g., Fischer and Brugger, 2011; Moeller et al., 2012; Wasner et al., in press). Many spatial-numerical associations are bodily experienced and might be represented in an embodied way, for instance, by using fingers for number magnitude. In recent intervention studies (Fischer et al., 2011; Link et al., 2013, 2014), it was shown that embodied spatial-numerical training leads to greater successful learning than various types of control training. Spatial experiences which are strongly routed in bodily representations may exert stronger influences on the build-up of SNAs, compared to other experiences. A similar account has been proposed by McCrink and Opfer (2014), who suggest that oriented motor behavior (e.g., hand movement during counting) might be a primary factor which refines SNAs in children. Following Fischer and Brugger (2011), one can postulate that for some SNAs embodied cultural influences like dominant reading/writing behavior may be most relevant (ordinality in counting), while for other SNAs (cardinality and its response side association) situated influences are more dominant.

We conclude that spatial-numerical directional preferences before formal reading should not be surprising. They need not be innate, because they may develop through many different cultural and social mechanisms. We suggest that their nature and consistency should be systematically studied. For future studies, we make several predictions:

- (i) Explicit SNAs (e.g., counting) should be trained best by explicit spatial-directional experiences, while implicit SNAs (e.g., SNARC) should be learned best through implicit spatial experiences.
- (ii) Conflicting spatial directions should lead to weaker directional SNAs than congruent spatial directions.
- (iii) Spatial learning mechanisms that are strongly embodied should influence SNAs more than mechanisms that are less strongly embodied or not embodied.

While these predictions are consistent with the available data, they have not been systematically tested so far. Future studies should not focus on the mere existence of different spatial-numerical associations in preschool children, but start exploring the relative contributions of distinct mechanisms which lead to the emergence and shape of distinct SNAs.

**ACKNOWLEDGMENTS**

We acknowledge support by Deutsche Forschungsgemeinschaft (DFG) and Open Access Publishing Fund of University of Tübingen for publishing open access. HN and UC were supported by the DFG grant CR-110/8-1.

**REFERENCES**

Bender, A., and Beller, S. (2012). Nature and culture of finger counting: diversity and representational effects of an embodied cognitive tool. *Cognition* 124, 156–182. doi: 10.1016/j.cognition.2012.05.005

Bloodgood, J. W. (1999). What’s in a name? Children’s name writing and literacy acquisition. *Read. Res. Q.* 34, 342–367. doi: 10.1598/RRQ.34.3.5

Briars, D., and Siegler, R. S. (1984). A featural analysis of preschoolers’ counting knowledge. *Dev. Psychol.* 20, 607–608. doi: 10.1037/0012-1649.20.4.607

- Chokron, S., and De Agostini, M. (1995). Reading habits and line bisection: a developmental approach. *Cogn. Brain Res.* 3, 51–58. doi: 10.1016/0926-6410(95)00018-6
- Chokron, S., and De Agostini, M. (2000). Reading habits influence aesthetic preference. *Cogn. Brain Res.* 10, 45–49. doi: 10.1016/S0926-6410(00)00021-5
- Chokron, S., and Imbert, M. (1993). Influence of reading habits on line bisection. *Cogn. Brain Res.* 1, 219–222. doi: 10.1016/0926-6410(93)90005-p
- Chokron, S., Kazandjian, S., and De Agostini, M. (2009). “Effects of reading direction on visuospatial organization: a critical review,” in *Quod Erat Demonstrandum: From Herodotus Ethnographic Journeys to Cross-Cultural Research*, eds A. Gari and K. Mylonas (Athens: Pedio Books Publishing), 107–114.
- Dehaene, S., Bossini, S., and Giraux, P. (1993). The mental representation of parity and number magnitude. *J. Exp. Psychol. Gen.* 122, 371–396. doi: 10.1037/0096-3445.122.3.371
- de Hevia, M. D., Girelli, L., Addabbo, M., and Cassia, V. M. (2014). Human infants’ preference for left-to-right oriented increasing numerical sequences. *PLoS ONE* 9:e96412. doi: 10.1371/journal.pone.0096412
- Doake, D. (1985). “Reading-like behavior: its role in learning to read,” in *Observing the Language Learner*, eds A. Jagger and M. Smith-Burke (Newark, DE: International Reading Association), 82–98.
- Dobel, C., Diesendruck, G., and Bölte, J. (2007). How writing system and age influence spatial representations of actions a developmental, cross-linguistic study. *Psychol. Sci.* 18, 487–491. doi: 10.1111/j.1467-9280.2007.01926.x
- Dooley, C. M. (2010). Young children’s approaches to books: the emergence of comprehension. *Read. Teach.* 64, 120–130. doi: 10.1598/RT.64.2.4
- Dyson, A. H. (1982). Reading, writing, and language: young children solving the written language puzzle. *Lang. Arts* 59, 829–839.
- Fagard, J., and Dahmen, R. (2003). The effects of reading-writing direction on the asymmetry of space perception and directional tendencies: a comparison between French and Tunisian children. *Laterality* 8, 39–52. doi: 10.1016/S0010-9452(08)70299-5
- Fischer, M. H. (2008). Finger counting habits modulate spatial-numerical associations. *Cortex* 44, 386–392. doi: 10.1016/j.cortex.2007.08.004
- Fischer, M. H., and Brugger, P. (2011). When digits help digits: spatial-numerical associations point to finger counting as prime example of embodied cognition. *Front. Psychol.* 2:260. doi: 10.3389/fpsyg.2011.00260
- Fischer, M. H., and Shaki, S. (2014). Spatial associations in numerical cognition – from single digits to arithmetic. *Q. J. Exp. Psychol.* 67, 1461–1483. doi: 10.1080/17470218.2014.927515
- Fischer, M. H., Shaki, S., and Cruise, A. (2009). It takes just one word to quash a SNARC. *Exp. Psychol.* 56, 361–366. doi: 10.1027/1618-3169.56.5.361
- Fischer, U., Moeller, K., Bientzle, M., Cress, U., and Nuerk, H. C. (2011). Sensorimotor spatial training of number magnitude representation. *Psychon. Bull. Rev.* 18, 177–183. doi: 10.3758/s13423-010-0031-3
- Fletcher, K. L., and Reese, E. (2005). Picture book reading with young children: a conceptual framework. *Dev. Rev.* 25, 64–103. doi: 10.1016/j.dr.2004.08.009
- Göbel, S. M., Fischer, M. H., and Shaki, S. (2014). “Experience counts: transmitting culture through space,” in *Poster and Oral Presentation at the 32. European Workshop on Cognitive Neuropsychology*, Bressanone.
- Güntürkün, O. (2003). Adult persistence of head-turning asymmetry. *Nature* 421, 711. doi: 10.1038/421711a
- Hildreth, G. (1936). Developmental sequences in name writing. *Child Dev.* 7, 291–303. doi: 10.2307/1125581
- Hoffmann, D., Hornung, C., Martin, R., and Schiltz, C. (2013). Developing number-space associations: SNARC effects using a colour-discrimination task in 5 year olds. *J. Exp. Child Psychol.* 116, 775–791. doi: 10.1016/j.jecp.2013.07.013
- Kamawar, D., LeFevre, J. A., Bisanz, J., Fast, L., Skwarchuk, S. L., Smith-Chant, B., et al. (2010). Knowledge of counting principles: how relevant is order irrelevance? *J. Exp. Child Psychol.* 105, 138–145. doi: 10.1016/j.jecp.2009.08.004
- Kazandjian, S., Cavézan, C., Zivotofsky, A. Z., and Chokron, S. (2010). Bisections in two languages: when number processing, spatial representation, and habitual reading direction interact. *Neuropsychologia* 48, 4031–4037. doi: 10.1016/j.neuropsychologia.2010.10.020
- Kazandjian, S., and Chokron, S. (2008). Paying attention to reading direction. *Nat. Rev. Neurosci.* 9, 965–965. doi: 10.1038/nrn2456-c1
- Kebbe, H., and Vinter, A. (2013). How culture, age, and manual dominance affect directionality in drawing side view objects. *J. Cross Cult. Psychol.* 44, 160–172. doi: 10.1177/0022022111435098
- Knudsen, B., Fischer, M. H., and Aschersleben, G. (in press). Development of spatial preferences for counting and picture naming. *Psychol. Res.* doi: 10.1007/s00426-014-0623-z
- Lieblich, A., Ninio, A., and Kugelmass, S. (1975). Developmental trends in directionality of drawing in Jewish and Arab Israeli children. *J. Cross Cult. Psychol.* 6, 504–511. doi: 10.1177/002202217564013
- Lindemann, O., Alipour, A., and Fischer, M. H. (2011). Finger counting habits in middle eastern and western individuals: an online survey. *J. Cross Cult. Psychol.* 42, 566–578. doi: 10.1177/0022022111406254
- Link, T., Moeller, K., Huber, S., Fischer, U., and Nuerk, H.-C. (2013). Walk the number line – an embodied training of numerical concepts. *Trends Neurosci. Educ.* 2, 74–84. doi: 10.1016/j.tine.2013.06.005
- Link, T., Schwarz, E. J., Huber, S., Fischer, U., Nuerk, H. C., Cress, U., et al. (2014). Mathe mit der Matte – Verkörperlichtes Training basisnumerischer Kompetenzen. *Z. Erziehungswiss.* 17, 257–277. doi: 10.1007/s11618-014-0533-532
- Maass, A., and Russo, A. (2003). Directional bias in the mental representation of spatial events nature or culture? *Psychol. Sci.* 14, 296–301. doi: 10.1111/1467-9280.14421
- Maass, A., Suitner, C., Favaretto, X., and Cignacchi, M. (2009). Groups in space: stereotypes and the spatial agency bias. *J. Exp. Soc. Psychol.* 45, 496–504. doi: 10.1016/j.jesp.2009.01.004
- McCrink, K., Birdsall, W., and Caldera, C. (2011). “Parental transmission of left-to-right spatial structuring in early childhood,” in *Poster Session Presented at the Seventh Biennial Meeting of the Cognitive Development Society in Philadelphia*, Philadelphia, PA.
- McCrink, K., and Opfer, J. E. (2014). Development of spatial-numerical associations. *Curr. Dir. Psychol. Sci.* 23, 439–445. doi: 10.1177/0963721414549751
- McCrink, K., Shaki, S., and Berkowitz, T. (2014). Culturally driven biases in preschoolers’ spatial search strategies for ordinal and non-ordinal dimensions. *Cogn. Dev.* 30, 1–14. doi: 10.1016/j.cogdev.2013.11.002
- Moeller, K., Fischer, U., Link, T. S., Wasner, M., Huber, S., Cress, U., et al. (2012). Learning and development of embodied numerosity. *Cogn. Process.* 13, 271–274. doi: 10.1007/s10339-012-0457-9
- Mol, S. E., Bus, A. G., and de Jong, M. T. (2009). Interactive book reading in early education: a tool to stimulate print knowledge as well as oral language. *Rev. Educ.* 79, 979–1007. doi: 10.3102/0034654309332561
- Nuerk, H. C., Iversen, W., and Willmes, K. (2004). Notational modulation of the SNARC and the MARC (linguistic markedness of response codes) effect. *Q. J. Exp. Psychol. A* 57, 835–863. doi: 10.1080/027249804343000512
- Opfer, J. E., and Furlong, E. E. (2011). How numbers bias preschoolers’ spatial search. *J. Cross Cult. Psychol.* 42, 682–695. doi: 10.1177/0022022111406098
- Opfer, J. E., Thompson, C. A., and Furlong, E. E. (2010). Early development of spatial numeric associations: evidence from spatial and quantitative performance of preschoolers. *Dev. Sci.* 13, 761–771. doi: 10.1111/j.1467-7687.2009.00934.x
- Patro, K., Fischer, U., Nuerk, H.-C., and Cress, U. (in press). How to rapidly construct a spatial-numerical representation in preliterate children (at least temporarily). *Dev. Sci.*
- Patro, K., and Haman, M. (2012). The spatial-numerical congruity effect in preschoolers. *J. Exp. Child Psychol.* 111, 534–542. doi: 10.1016/j.jecp.2011.09.006
- Patro, K., Nuerk, H. C., Cress, U., and Haman, M. (2014). How number-space relationships are assessed before formal schooling: a taxonomy proposal. *Front. Psychol.* 5:419. doi: 10.3389/fpsyg.2014.00419
- Puranik, C. S., Lonigan, C. J., and Kim, Y. S. (2011). Contributions of emergent literacy skills to name writing, letter writing, and spelling in preschool children. *Early Child. Res. Q.* 26, 465–474. doi: 10.1016/j.ecresq.2011.03.002
- Puranik, C. S., Schreiber, S., Estabrook, E., and O’Donnell, E. (2013). Comparison of name-writing rubrics: is there a gold standard? *Assess. Eff. Interv.* 40, 16–23. doi: 10.1177/1534508413502390
- Rinaldi, L., Di Luca, S., Henik, A., and Girelli, L. (2014). Reading direction shifts visuospatial attention: an interactive account of attentional biases. *Acta Psychol.* 151, 98–105. doi: 10.1016/j.actpsy.2014.05.018
- Rugani, R., Vallortigara, G., Priftis, K., and Regolin, L. (2015). Newborn chicks map numbers to space similarly to humans. *Science* 347, 534–536.

- Rugani, R., Vallortigara, G., and Regolin, L. (in press). At the root of the left–right asymmetries in spatial–numerical processing: from domestic chicks to human subjects. *J. Cogn. Psychol.* doi: 10.1080/20445911.2014.941846
- Sénéchal, M., Cornell, E. H., and Broda, L. S. (1995). Age-related differences in the organization of parent–infant interactions during picture-book reading. *Early Child. Res. Q.* 10, 317–337. doi: 10.1016/0885-2006(95)90010-1
- Sénéchal, M., LeFevre, J. A., Hudson, E., and Lawson, E. P. (1996). Knowledge of storybooks as a predictor of young children’s vocabulary. *J. Educ. Psychol.* 88, 520. doi: 10.1037/0022-0663.88.3.520
- Shaki, S. (2013). What’s in a kiss? Spatial experience shapes directional bias during kissing. *J. Nonverbal Behav.* 37, 43–50. doi: 10.1007/s10919-012-0141-x
- Shaki, S., and Fischer, M. H. (2008). Reading space into numbers – a crosslinguistic comparison of the SNARC effect. *Cognition* 108, 590–599. doi: 10.1016/j.cognition.2008.04.001
- Shaki, S., Fischer, M., and Göbel, S. (2012). Direction counts: a comparative study of spatially directional counting biases in cultures with different reading directions. *J. Exp. Child Psychol.* 112, 275–281. doi: 10.1016/j.jecp.2011.12.005
- Shaki, S., Fischer, M. H., and Petrusic, W. M. (2009). Reading habits for both words and numbers contribute to the SNARC effect. *Psychon. Bull. Rev.* 16, 328–331. doi: 10.3758/pbr.16.2.328
- Snow, C., Burns, S., and Griffin, P. (1998). *Preventing Reading Difficulties in Young Children*. Washington, DC: The National Research Council.
- Spalek, T. M., and Hammad, S. (2005). The left-to-right bias in inhibition of return is due to the direction of reading. *Psychol. Sci.* 16, 15–18. doi: 10.1111/j.0956-7976.2005.00774.x
- Sulzby, E. (1985). Children’s emergent reading of favorite storybooks: a developmental study. *Read. Res. Q.* 20, 458–481. doi: 10.1598/RRQ.20.4.4
- Tversky, B., Kugelmass, S., and Winter, A. (1991). Cross-cultural and developmental trends in graphic productions. *Cogn. Psychol.* 23, 515–557. doi: 10.1016/0010-0285(91)90005-9
- Vaid, J. (1995). “Script directionality affects nonlinguistic performance: evidence from Hindi and Urdu,” in *Scripts and Literacy*, eds I. Taylor and D. R. Olson (Amsterdam: Springer), 295–310.
- Valencia, S. W., and Sulzby, E. (1991). Assessment: assessment of emergent literacy: storybook reading. *Read. Teach.* 44, 498–500.
- van Dijck, J. P., and Fias, W. (2011). A working memory account for spatial–numerical associations. *Cognition* 119, 114–119. doi: 10.1016/j.cognition.2010.12.013
- Wasner, M., Moeller, K., Fischer, M. H., and Nuerk, H.-C. (in press). Related but not the same: ordinality, cardinality and 1-to-1 correspondence in finger-based numerical representations. *J. Cogn. Psychol.* doi: 10.1080/20445911.2014.964719
- Wood, G., Willmes, K., Nuerk, H. C., and Fischer, M. H. (2008). On the cognitive link between space and number: a meta-analysis of the SNARC effect. *Psychol. Sci. Q.* 50, 489–525.
- Zebian, S. (2005). Linkages between number concepts, spatial thinking, and directionality of writing: the SNARC effect and the REVERSE SNARC effect in English and Arabic monoliterates, biliterates, and illiterate Arabic speakers. *J. Cogn. Cult.* 5, 165–190. doi: 10.1163/1568537054068660

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 14 November 2014; paper pending published: 21 December 2014; accepted: 11 February 2015; published online: 05 March 2015.

Citation: Nuerk H-C, Patro K, Cress U, Schild U, Friedrich CK and Göbel SM (2015) How space-number associations may be created in preliterate children: six distinct mechanisms. *Front. Psychol.* 6:215. doi: 10.3389/fpsyg.2015.00215

This article was submitted to *Developmental Psychology*, a section of the journal *Frontiers in Psychology*.

Copyright © 2015 Nuerk, Patro, Cress, Schild, Friedrich and Göbel. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Up or down? Reading direction influences vertical counting direction in the horizontal plane – a cross-cultural comparison

**Silke M. Göbel\***

Department of Psychology, University of York, York, UK

**Edited by:**

Hans-Christoph Nuerk, University of Tuebingen, Germany

**Reviewed by:**

Maria Dolores De Hevia, Centre National de la Recherche Scientifique, France

Matthias Hartmann, University of Potsdam, Germany

Martin Lachmair, Leibniz-Institut für Wissensmedien, Germany

**\*Correspondence:**

Silke M. Göbel, Department of Psychology, University of York, Heslington, York, YO10 5DD, UK  
e-mail: silke.goebel@york.ac.uk

Most adults and children in cultures where reading text progresses from left to right also count objects from the left to the right side of space. The reverse is found in cultures with a right-to-left reading direction. The current set of experiments investigated whether vertical counting in the horizontal plane is also influenced by reading direction. Participants were either from a left-to-right reading culture (UK) or from a mixed (left-to-right and top-to-bottom) reading culture (Hong Kong). In Experiment 1, native English-speaking children and adults and native Cantonese-speaking children and adults performed three object counting tasks. Objects were presented flat on a table in a horizontal, vertical, and square display. Independent of culture, the horizontal array was mostly counted from left to right. While the majority of English-speaking children counted the vertical display from bottom to top, the majority of the Cantonese-speaking children as well as both Cantonese- and English-speaking adults counted the vertical display from top to bottom. This pattern was replicated in the counting pattern for squares: all groups except the English-speaking children started counting with the top left coin. In Experiment 2, Cantonese-speaking adults counted a square array of objects after they read a text presented to them either in left-to-right or in top-to-bottom reading direction. Most Cantonese-speaking adults started counting the array by moving horizontally from left to right. However, significantly more Cantonese-speaking adults started counting with a top-to-bottom movement after reading the text presented in a top-to-bottom reading direction than in a left-to-right reading direction. Our results show clearly that vertical counting in the horizontal plane is influenced by longstanding as well as more recent experience of reading direction.

**Keywords:** mental number line, grounded cognition, SNARC, spatial-numerical association, children, physical world account

## INTRODUCTION

Spoken language affects various aspects of number processing and arithmetic. For example, the way number words are constructed differs between languages. The complexity of number word construction influences early counting, arithmetic and place-value understanding (Dowker et al., 2008; Siegler and Mu, 2008; Zuber et al., 2009) and inconsistencies between the Arabic notation and number word construction (e.g., number word inversion) lead to disadvantages in symbolic number processing (Pixner et al., 2011) and affects symbolic arithmetic (Göbel et al., 2014b). Written language practices also affect numerical cognition. For example, the direction of reading and writing within a culture can influence number processing (Göbel et al., 2011). The current paper focuses on the influence of reading direction on the direction of counting by comparing the counting of children and adults in the United Kingdom (UK) to children and adults from Hong Kong (HK).

Most Western adults and children count objects horizontally from left to right (Opfer and Thompson, 2006; Opfer et al., 2010; Shaki et al., 2012). This counting bias might be yet another instantiation of the mental number line, a common spatial-numerical association (SNA) of small numbers with left

and larger numbers with right space (Fischer and Brugger, 2011). Evidence for a mental number line with a left-to-right direction comes from a large body of research investigating the spatial-numerical association of response codes (SNARC) effect: in parity judgment participants are consistently faster to respond with left responses to smaller and right responses to larger numbers (Dehaene et al., 1993; for a review see Wood et al., 2008).

Interestingly, several recent studies have reported the existence of horizontal SNAs already in young infants and animals: newly hatched and 3-day-old chicks have a tendency to associate large numbers with the right side of space (Rugani et al., 2014, 2015), chimpanzees and rhesus monkeys associate smaller numbers with starting on the left side of space (Adachi, 2014; Drucker and Brannon, 2014) and 7-month-old infants prefer displays that increase in magnitude to be shown from left to right (de Hevia et al., 2014). These findings point toward a biological predisposition for early horizontal SNAs (Rugani et al., 2010, 2011). Hemispheric lateralization could account for an advantage in processing the left hemisphere: an early right hemispheric dominance in visuo-spatial tasks might lead to a stronger allocation of attention to the left hemifield (Mesulam, 1990). Combined with a preference for

increasing sequences (Macchi Cassia et al., 2012) this hemispheric asymmetry could provide the early building blocks of a left-to-right SNA. Some evidence for this *hemispheric asymmetry account* comes from adult neglect patients (Heilman and Van Den Abell, 1980): after damage to their right parietal lobe they typically show a rightward shift in line and number bisection (Umiltà et al., 2009), pointing toward a role of the right parietal lobe in attending toward left space (Göbel et al., 2006). However, the hemispheric asymmetry account cannot explain why illiterate adults showed no significant SNARC effect (Zebian, 2005) and no preference for a particular horizontal counting direction (Shaki et al., 2012). Studies on illiterate adults provide strong support for an alternative account for horizontal SNAs: *the reading direction account*.

This account suggests that the direction of the mental number line is shaped by the culturally dominant reading direction. Already in the first paper on the SNARC effect Dehaene et al. (1993) provided evidence that the size and possibly the direction of this effect might be related to reading direction. They investigated the SNARC effect in a group of participants who originated from a right-to-left reading culture (Iran) but were living in a left-to-right reading culture (France) at the time of testing. The strength of their SNARC effect was correlated with the length of time spent in the left-to-right reading culture. In addition, there is evidence for a reversal in the direction of the SNARC effect and the dominant counting direction in cultures with a right-to-left reading direction. Arab participants who read from right to left show a reversed SNARC effect: they are faster to respond to small numbers with a right and to larger numbers with a left response (Zebian, 2005; Shaki et al., 2009). Similarly, the majority of Arab adults and children count from right to left (Shaki et al., 2012). In summary, those findings are most convincingly explained by the reading direction account.

Taking this account a step further, the reading direction account predicts a vertical mental number line in cultures reading from top to bottom. At this point it is important to clarify that the term vertical is used in two ways: in a two-dimensional context, for example when reading a page of a book, the vertical axis refers to the axis perpendicular to the horizontal axis. However, in 3D the true vertical axis is perpendicular to the horizontal plane. Surprisingly little research has investigated SNAs in the vertical dimension and most research on vertical SNAs so far has focused on the vertical axis in the horizontal plane (see Hartmann et al., 2014).

During number processing some people automatically activate visuo-spatial images of number lines (so called number forms) that are stable over time and highly individual. Already in an early description, many of these forms (Galton, 1880, Figures 2, 4, 6 and 8) progress not only from left to right but also from bottom to top. In a study of 15 Belgium university students with number forms, nine number forms progressed from the bottom up and only one from top to bottom (Seron et al., 1992). Sagiv et al. (2006) classified the direction of number forms of 114 Scottish synaesthetes and 311 controls without synaesthesia as either left-to-right, right-to-left, bottom-to-top, or top-to bottom exclusively. The majority was classified as left-to-right, but 11% of the synaesthetes' number forms and 23% of the controls' number forms progressed bottom-to-top and none showed a top-to-bottom direction. This suggests, at least in individuals with number forms, a predominant vertical

association of small numbers with bottom and larger numbers with top space.

Research suggests that this vertical association is not specific to just people with explicit number forms. Schwarz and Keus (2004) found a truly vertical SNARC effect in Dutch participants: eye movements to a bottom response location started earlier for smaller than larger numbers while eye movements to a top response location begun earlier for larger than smaller numbers. Further, a vertical SNARC effect has been found in Belgium, American, German, and Israeli participants (Gevers et al., 2006; Müller and Schwarz, 2007; Holmes and Lourenco, 2012; Shaki and Fischer, 2012; Hartmann et al., 2014). The majority of these studies (Gevers et al., 2006; Müller and Schwarz, 2007; Shaki and Fischer, 2012) used vertical responses in the horizontal plane, i.e., close and far response buttons. However, two of these studies (Holmes and Lourenco, 2012; Hartmann et al., 2014) used a truly vertical response button arrangement and found that participants were faster to respond to small numbers with bottom hand responses and large numbers with top hand responses.

Those findings support the idea of a vertical dimension of number magnitude with increasing magnitude from bottom to top. This direction of the vertical SNA is opposite to predictions from the reading direction account. At first, one might think that in Western participants the dominant reading direction is from left to right and thus neutral with respect to the vertical dimension. However, given that most reading and writing in adults involves more than one line of text, reading and writing have a secondary direction: line by line, from the top to the bottom of a page. A strong version of the reading direction account thus proposes that the secondary reading direction (top-to-bottom) should also influence the direction of the SNA and lead to an association of small numbers with top and larger numbers with bottom space. However, I suggest an alternative: a weaker version of the reading direction account proposes that only the dominant reading direction is affecting SNAs and not the secondary reading direction. This weaker version can account for the horizontal SNA, but is silent with respect to the vertical SNA found in left-to-right and right-to-left reading cultures. Interestingly, this hints at possibly different mechanisms underlying horizontal and vertical SNAs.

Vertical associations might reflect experience with the physical world (Lakoff and Núñez, 2000; Gevers et al., 2006). In the physical world magnitude is often associated with higher up in the vertical dimension: more water in a glass is indicated by a higher level, higher buildings and trees and taller people extend more upward than smaller ones. If the association between number magnitude and vertical space is mainly driven by experiences in the physical world (*the physical world account*) then the association of small numbers with the bottom and larger numbers with the top space should be found independent of cultural context. In Fischer's (2012) terminology, this physical world account is a grounded theory (Barsalou, 2008, p. 162) based on "invariants in the physical world". Support for this account comes for example from a study by Lachmair et al. (2014). In a lexical decision task, after being primed with small numbers, participants were significantly faster to respond to words that are normally associated with lower vertical space (e.g., submarine). In contrast, words

associated with upper vertical space (e.g., eagle) were significantly faster responded to when the prime was a large number.

Research on vertical SNAs in Japan, a culture with a dominant reading direction from top to bottom, strongly supports the physical world account. Ito and Hatta (2004) asked 50 Japanese undergraduate students to place 0–9 on a vertical line. The majority (76%) placed ascending numbers from bottom to top and only 18% used a top-to-bottom arrangement, arguing against a dominant influence of vertical reading direction. When Japanese participants performed a vertical SNARC task with response buttons in the horizontal plane they also showed the same association as Western participants: Japanese participants were faster to respond to smaller numbers with bottom than top responses and to larger numbers with top than bottom responses. The direction of their vertical SNAs was opposite to their reading direction and in line with the physical world account.

A study with Taiwanese participants, however, showed that whether reading direction influences the SNARC effect in the horizontal and vertical dimension might depend on the number format used in the task. There are three numerical notations in Taiwan: (1) Arabic digits (e.g., 1), (2) Chinese number words in the simple form (e.g., 一), (3) Chinese number words in the complex form (e.g., 壹). Hung et al. (2008) tested the horizontal and vertical SNAs of these three notations in Taiwanese participants. Arabic digits are typically printed horizontally in text in Taiwan, while Chinese number words appear more often in vertical text with a top-to-bottom directionality. For Arabic digits they found a significant horizontal SNARC effect with faster left than right responses for smaller digits and faster right than left responses for larger digits, but there was no significant horizontal SNARC effect for Chinese number words. In contrast, the vertical association between numbers and space was only significant for the Chinese number words in the simple form, but not for Arabic digits or Chinese number words in the complex form. Chinese number words in the simple form were responded to faster with top than bottom responses for small numbers and faster with bottom than top responses for large numbers. This suggests that the association between number and space is not hardwired, but flexible (Bächtold et al., 1998; Ristic et al., 2006; Fischer et al., 2009, 2010) and can adapt rapidly to a different context. In Fischer's (2012) terminology, this speaks for the situatedness of SNAs. Furthermore, it was the dominant reading direction associated with the specific number notation used in the task that predicted the specific direction of the SNA. The different results found for Chinese number words in the simple and in the complex form suggest that in order to influence SNAs the association between notation and reading direction needs to be strong and firmly established. Chinese number words in the complex form probably did not influence the direction of SNA, because they are less frequent than Arabic digits and Chinese number words in the simple form and do not strongly evoke a reading context.

In summary, SNAs also exist in the vertical dimension. The most common association seems to be along a mental number line with numbers with increasing magnitude going from bottom to top space. Reading direction possibly can influence this association under certain conditions.

The first aim of the current study was to investigate whether reading direction influences the direction of vertical counting in the horizontal plane. To the best of my knowledge, this has not yet been investigated. An explicitly spatial-numerical task (object counting) was chosen rather than the implicit, more commonly used SNA task of number judgment because we have shown that reading direction influences the horizontal counting direction (Shaki et al., 2012). Furthermore, so far no study has directly investigated the effect of reading direction on implicit SNA tasks in young children while there is evidence from our own work (Göbel et al., 2014a) that recent reading observation, even in preliterate children, can change their horizontal counting direction. Investigating vertical counting was logically the next step. I chose two groups of participants with different reading experiences: (1) participants with a dominant reading direction from left to right and a secondary reading direction from top to bottom (UK), (2) participants with mixed dominant and secondary reading direction (Hong Kong [HK]). The majority of text in books and newspapers in HK is printed from left to right with a secondary reading direction from top to bottom. A visible minority of text, however, is presented in top-to-bottom direction with the secondary reading direction going from right to left<sup>1</sup>. I was interested in the effect of both dominant and secondary reading direction on the direction of counting. The second aim was to investigate whether the amount of reading (and writing) experience influences the strength of the association. We therefore tested both children and adults. Children were beginning readers and had thus much smaller experience with the cultural direction of reading and writing than adults. Third, given that there might be different mechanisms underlying horizontal and vertical SNAs I was interested in whether there is a hierarchy in the association of number and space. For example, are horizontal SNAs more dominant than vertical SNAs? We tested this by asking participants to count objects in a display with balanced vertical and horizontal dimensions (a square of objects). Lastly, we were interested in how flexible those spatial biases are. Thus, in Experiment 2 we manipulated the most recent reading experience direction (left-to-right or top-to-bottom) and investigated whether the most recent reading experience shows an immediate effect on counting direction.

## EXPERIMENT 1

Adults and children in the UK and in HK were asked to count objects in three differently arranged displays: a horizontal, a vertical, and a square display (**Figure 1**). In line with their dominant reading direction, we expected the majority of all participants to count the horizontal array from left to right. With respect to counting the vertical array, the strong reading direction account predicts that all participants will count from top to bottom, while the weak reading account predicts no preference for a specific vertical counting direction in UK participants, but a top-to-bottom preference

<sup>1</sup> Many street and shop signs in HK are vertical. In a pilot study in 2012, 100 books for children and 100 books for adults were picked randomly from the shelves of the Hong Kong Central Library. Eighty percent of the books for children used a left-to-right reading direction, 17% a top-to-bottom reading direction, and 3% mixed reading directions. For adult books the corresponding percentages were: 46% left-to-right, 42% top-to-bottom, and 12% mixed.

for HK participants. We expected the children to show this pattern less strongly than the adults due to their limited experience with reading and writing. The physical world account, in contrast, predicts that most participants will count the vertical array from bottom to top. For counting objects in the square arrangement there are two factors of interest: first, the starting position and second, the direction of the first movement. The reading account predicts a top left starting position and a first movement from left to right for all participants. The physical world account predicts a bottom starting position and a first movement from bottom to top, but is neutral with respect to left or right side.

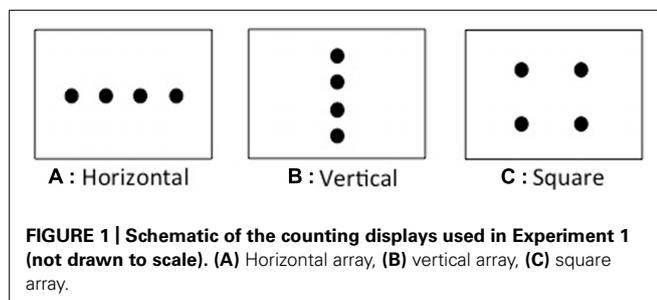
## MATERIALS AND METHODS

### Participants

All British participants (80 children, 100 adults) were native English speakers brought up in the UK. All HK-Chinese participants (94 children, 99 adults) were native Cantonese speakers brought up in HK. British 4- and 5-year-old children were tested with parental consent in nurseries and primary schools in North Yorkshire, Greater Manchester, and Shrewsbury. HK-Chinese 4- and 5-year-old children were tested with parental consent in kindergartens in HK. All adult participants gave written consent. British adults were tested in the UK, HK-Chinese adults in HK. Data for left-handed children and adults were excluded. I am reporting data for the remaining 71 British children (mean age = 4.44 years, SD = 0.50, 33 female, 38 male), 85 HK-Chinese children (mean age = 4.82, SD = 0.38, 51 female, 34 male), 90 British adults (18–94 years, mean age = 48.07 years, SD = 21.64, 58 female, 32 male) and 99 HK-Chinese adults (18–83 years, mean age = 32.98, SD = 14.80, 59 female, 40 male). The study was approved by the Ethics Committee, Department of Psychology, University of York.

### Materials

Twelve golden plastic coins (diameter = 3.5 cm) and three rectangular mats (40 cm × 30 cm, landscape) were used to create three counting displays (Figure 1). For the horizontal display four coins were placed horizontally in a linear array onto the mat, equidistant (4.0 cm) from each other with the two outer coins placed at about 6.3 cm from the side edges of the mat and all coins at about 13.3 cm from the top and bottom edges of the mat. In the vertical display four coins were placed flat on the mat, vertically in a linear array equidistant (3.0 cm) from each other with the coins placed at about 18.3 cm from the side edges and at about 3.5 cm from the top and bottom edges. In the square display four coins were placed into a 2 by 2 square, with about 8.0 cm between each coin,



with the outer edges of the square arrangement at about 12.5 cm from the left and right edges of the mat and at about 7.5 cm from the top and bottom edges.

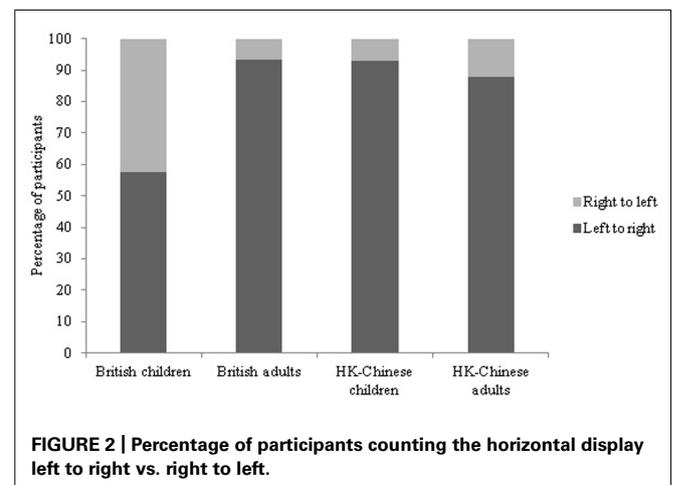
### Procedure

All three stimuli sets (horizontal, vertical, and square display) were prepared before testing and covered with DIN A3 sheets of paper. Participants were tested individually in a quiet room. HK-Chinese participants were tested in Cantonese, by a native Cantonese speaker. British participants were tested in English, by a native English speaker. Stimuli were present lying flat on the table at which the participant was seated, centrally in front of the participant, and covered. The first stimulus set was then presented by lifting off the cover. Participants were asked, “Can you please point to each of the coins for me and count aloud how many there are?” No demonstration was given, and participants’ order of counting was recorded by the experimenter. The instruction was repeated twice again with the next two stimulus sets. Next, handedness was tested. Children were asked to draw a picture of a sun. Adults filled out the Edinburgh Handedness Questionnaire (Oldfield, 1971). In addition, children in HK were asked to write three age-appropriate characters (big, small, mother). At the end participants were thanked, children were praised, and received a sticker. All participants counted the horizontal, vertical, and square displays. The order of the presentation of the three displays and the seating position of the experimenter (to the left or right of the participant) was counterbalanced between participants.

## RESULTS

### Horizontal array

As can be seen in Figure 2, the majority of all participants counted the horizontal display from left to right (57.7% of the British children, 93.3% of the British adults, 92.9% of the HK-Chinese children, and 87.9% of the Chinese adults, Supplementary Table S1). The difference between the number of participants counting left to right and right to left was significant for British adults ( $\chi^2 = 67.6$ ,  $df = 1$ ,  $p < 0.01$ ), HK-Chinese children ( $\chi^2 = 62.69$ ,  $df = 1$ ,  $p < 0.01$ ) and HK-Chinese adults ( $\chi^2 = 56.82$ ,  $df = 1$ ,  $p < 0.01$ ). For the British children, there was no significant



preference in counting direction ( $\chi^2 = 1.70$ ,  $df = 1$ ,  $p = 0.19$ ). Significantly more British adults counted from left to right than British children ( $\chi^2 = 29.0$ ,  $df = 1$ ,  $p < 0.001$ ). There was no significant difference in counting frequency between the HK-Chinese children and adults ( $\chi^2 = 1.33$ ,  $df = 1$ ,  $p = 0.25$ ) or between the British adults and HK-Chinese adults ( $\chi^2 = 1.63$ ,  $df = 1$ ,  $p = 0.20$ ). Although more British 5-year-olds (67.7%) than 4-year-olds (50.0%) counted left to right this difference did not reach significance ( $\chi^2 = 2.25$ ,  $df = 1$ ,  $p = 0.13$ ) and there was no effect of age on the horizontal counting direction for HK-Chinese children either ( $\chi^2 = 1.38$ ,  $df = 1$ ,  $p = 0.24$ ). There was no effect of order, experimenter location or gender on the frequency of horizontal counting direction in any group (all  $ps > 0.05$ ).

### Vertical array

Significantly more British children (74.6%) counted from bottom to top than from top to bottom (25.4%;  $\chi^2 = 17.25$ ,  $df = 1$ ,  $p < 0.01$ ). In contrast, the majority of British adults (83.3%,  $\chi^2 = 40.00$ ,  $df = 1$ ,  $p < 0.01$ ), HK-Chinese children (81.2%,  $\chi^2 = 33.05$ ,  $df = 1$ ,  $p < 0.01$ ) and HK-Chinese adults (86.9%,  $\chi^2 = 53.83$ ,  $df = 1$ ,  $p < 0.01$ ) counted from top to bottom (see **Figure 3**, Supplementary Table S1). The counting patterns between British children and adults were significantly different ( $\chi^2 = 54.7$ ,  $df = 1$ ,  $p < 0.01$ ). There was no significant difference in counting frequency neither between the HK-Chinese children and adults ( $\chi^2 = 1.12$ ,  $df = 1$ ,  $p = 0.29$ ) nor between the British adults and the HK-Chinese adults ( $\chi^2 = 0.47$ ,  $df = 1$ ,  $p = 0.49$ ), but there was a significant difference between HK-Chinese children and British children ( $\chi^2 = 48.90$ ,  $df = 1$ ,  $p < 0.001$ ). There was no significant difference in counting preference between the 4- and 5-year-old children neither for the British children nor for the HK-Chinese children (all  $ps > 0.05$ ). Gender did not affect the counting direction (all  $ps > 0.05$ ). While there was no effect of order or experimenter location for British or Chinese adults (all  $ps > 0.05$ ), order had a significant effect for both British, and HK-Chinese children. For British children significantly more children counted top to bottom when the vertical array came after the square (50.0%) than when it came before the square array (15.75%;  $\chi^2 = 8.94$ ,  $df = 1$ ,  $p < 0.001$ ). The same pattern was observed for the HK-Chinese children: significantly more Chinese

children counted top to bottom when the vertical array came after the square array (90.24%) than when it came before the square array (72.72%;  $\chi^2 = 4.26$ ,  $df = 1$ ,  $p = 0.04$ ). Experimenter location did not affect counting direction for the HK-Chinese children ( $\chi^2 = 0.253$ ,  $df = 1$ ,  $p = 0.62$ ). However, significantly more British children with the experimenter sitting on their right side (36.3%) counted top to bottom than British children with the experimenter sitting on their left side (15.8%,  $\chi^2 = 3.95$ ,  $df = 1$ ,  $p = 0.047$ ).

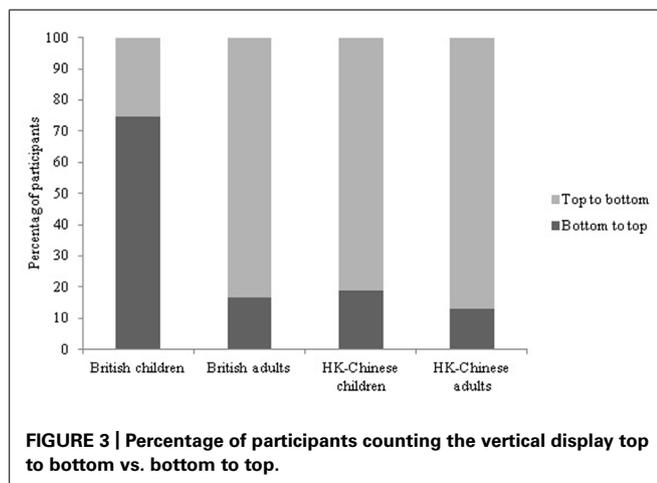
### Square array

The data for one 5-year-old British child were excluded from the data analysis because he moved diagonally when counting the coins in the square. The majority of the British children started to count either on the bottom left (31.4%) or right (41.4%) coin. All other groups showed a clear preference to start counting with the top left coin (British adults: 88.9%, HK-Chinese children: 71.8%; HK-Chinese adults: 81.8%; see **Table 1**).

**Vertical starting position.** Significantly more British children (72.9%) started counting the square on one of the two bottom coins than on one of the two top coins (27.1%;  $\chi^2 = 14.63$ ,  $df = 1$ ,  $p < 0.01$ ). In contrast, the majority of British adults (92.2%,  $\chi^2 = 64.18$ ,  $df = 1$ ,  $p < 0.01$ ), HK-Chinese children (80.0%,  $\chi^2 = 30.60$ ,  $df = 1$ ,  $p < 0.01$ ) and HK-Chinese adults (93.9%,  $\chi^2 = 76.46$ ,  $df = 1$ ,  $p < 0.01$ ) started at a top coin. The counting patterns between British children and British adults ( $\chi^2 = 72.2$ ,  $df = 1$ ,  $p < 0.01$ ) as well as between the British children and the HK-Chinese children ( $\chi^2 = 43.6$ ,  $df = 1$ ,  $p < 0.001$ ) were significantly different. Although overall most HK-Chinese children and adults counted the square starting from a top coin, there were significantly more HK-Chinese children (20.0%) who started counting the coins in the square from a bottom coin than HK-Chinese adults (6.1%,  $\chi^2 = 8.12$ ,  $df = 1$ ,  $p < 0.01$ ). There was no significant difference in counting preference between the British adults and the HK-Chinese adults ( $\chi^2 = 0.22$ ,  $df = 1$ ,  $p = 0.64$ ).

**Horizontal starting position.** All groups except the British children showed a clear preference to start counting the coins in the square on the left side (British adults: 91.1%,  $\chi^2 = 60.84$ ,  $df = 1$ ,  $p < 0.01$ ; HK-Chinese children: 80.0%,  $\chi^2 = 30.60$ ,  $df = 1$ ,  $p < 0.01$ , HK-Chinese adults: 84.8%,  $\chi^2 = 49.5$ ,  $df = 1$ ,  $p < 0.01$ ). For British children there was no significant difference between the number of children starting counting the coins in the square on the left (50.0%) versus on the right side (50.0%,  $\chi^2 = 0$ ,  $df = 1$ ,  $p = 1.00$ ). The counting patterns between British children and British adults ( $\chi^2 = 33.9$ ,  $df = 1$ ,  $p < 0.001$ ) as well as between the British children and the HK-Chinese children ( $\chi^2 = 15.5$ ,  $df = 1$ ,  $p < 0.001$ ) were significantly different. There was no significant difference in counting preference neither between the HK-Chinese children and HK-Chinese adults ( $\chi^2 = 0.748$ ,  $df = 1$ ,  $p = 0.39$ ) nor between British adults and the HK-Chinese adults ( $\chi^2 = 1.73$ ,  $df = 1$ ,  $p = 0.19$ ).

**First movement.** As expected, the first movement when counting the coins in the square was horizontal for most British adults (88.9%,  $\chi^2 = 54.4$ ,  $df = 1$ ,  $p < 0.01$ ), HK-Chinese children (69.4%,



**Table 1 | Number of participants by starting position and direction of first movement for counting the square display for Experiment 1.**

Group	Starting position				First movement			
	Left		Right		Horizontal		Vertical	
	Top	Bottom	Top	Bottom	Left-right	Right-left	Bottom-top	Top-bottom
<b>British</b>								
Children	13	22	6	29	24	15	26	5
Adults	80	2	3	5	77	3	4	6
<b>HK-Chinese</b>								
Children	61	7	7	10	54	5	11	15
Adults	81	3	12	3	76	4	2	17

$\chi^2 = 12.81$ ,  $df = 1$ ,  $p < 0.01$ ) and HK-Chinese adults (80.8%,  $\chi^2 = 37.59$ ,  $df = 1$ ,  $p < 0.01$ ) and the majority moved from left to right (British adults: 85.6%, HK-Chinese children: 63.5%, HK-Chinese adults: 76.8%). In contrast, for British children there was no significant preference for moving horizontally (55.7%) or vertically (44.2%;  $\chi^2 = 0.914$ ,  $df = 1$ ,  $p = 0.34$ ) first. This pattern was significantly different from the counting pattern for British adults ( $\chi^2 = 22.7$ ,  $df = 1$ ,  $p < 0.001$ ) and for HK-Chinese adults ( $\chi^2 = 12.4$ ,  $df = 1$ ,  $p < 0.001$ ) and approaching a significant difference to the counting pattern for HK-Chinese children ( $\chi^2 = 3.10$ ,  $df = 1$ ,  $p = 0.08$ ). 34.3% of British children's first movement was from left to right, 21.4% from right to left, 37.2% from bottom to top and 7.1% from top to bottom. For more details, please see **Table 1**.

#### **Experimenter seating position, order, gender, and children's age.**

There were no significant differences between the square counting patterns of 4 and 5 year olds for the British or the HK-Chinese children and no effect of gender (all  $ps > 0.05$ ). For British children and adults as well as for HK-Chinese adults experimenter seating position and order of the square array did not significantly affect their counting behavior (all  $ps > 0.05$ ). However, for the HK-Chinese children significantly more children (29.5%) started counting at the bottom than the top when the square was presented after the vertical display than when it was presented before (9.8%,  $\chi^2 = 5.19$ ,  $df = 1$ ,  $p = 0.02$ ). Equally, their first movement was significantly more likely to be vertical when the square display was presented after the vertical array (40.9%) than when it was presented before the vertical array (19.5%,  $\chi^2 = 4.58$ ,  $df = 1$ ,  $p = 0.03$ ). In addition, significantly more HK-Chinese children started counting the square on the right side when the experimenter was sitting on their left (32.6%) than when she was sitting on their right side (7.7%,  $\chi^2 = 8.58$ ,  $df = 1$ ,  $p < 0.01$ ). All other effects of order and experimenter location were non-significant.

#### **DISCUSSION**

Overall, results from Experiment 1 broadly support the reading direction account. As predicted, there was a preference to count the horizontal array from left to right. For British children this preference was present, but not statistically significant.

For all other groups the left-to-right preference was statistically significant supporting the reading direction account of horizontal SNAs. At first, the finding that British children did not show a significant preference of horizontal counting direction seems to be at odds with previous findings of a left-to-right SNA in 3-6-year-old Western children (Patro and Haman, 2012; Shaki et al., 2012; Knudsen et al., in press). However, previous studies have shown that horizontal SNAs can only be elicited in young children under certain conditions (Hoffmann et al., 2013) and that they are less pronounced than in older children or even absent (Berch et al., 1999; Van Galen and Reitsma, 2008). In addition, the percentage of children counting left to right found in our study (57.7%) is comparable to a previous study in which 60.7% of UK pre-school children showed a preference for counting from left to right (Shaki et al., 2012). In this study the preference for counting from left to right in UK children increased significantly from preschool into school age lending further support to the reading direction account.

For the vertical array, in line with the strong reading direction account, the majority of adults and HK-Chinese children in our study counted the coins from the top to the bottom. British children, however, showed a significant preference to count from bottom to top. A similar pattern was observed for counting the square array: while most British adults, HK-Chinese adults and HK-Chinese children started counting with the top left coin, British children preferred to start counting with a bottom coin with no preference for either the left or right bottom coin. In summary, the reading direction account explains the findings from British adults, HK-Chinese adults, and HK-Chinese children.

In contrast, the counting patterns of British children are in line with the physical world account. Although we chose the same age groups for British and HK-Chinese children, HK-Chinese children start being taught to write (and read) in Chinese from around age 3 (Curriculum Development Council, 2006). This is much earlier than for British children. I suggest that the differences in counting patterns between the British and the HK-Chinese children are explained by the fact that the two groups of children were not matched on reading (and writing) experience. For young children with little reading skill the experience of magnitude in the physical world might dominate their vertical SNAs and the culturally

dominant reading direction only begins to shape their SNAs with increasing exposure to and experience of reading and writing. There are two aspects of our data that support this conclusion: first, although HK-Chinese children showed a clear preference to count the square starting at the top left coin, significantly more HK-Chinese children than HK-Chinese adults started counting from a bottom coin, showing some residual pattern in line with the physical world account. Second, although British children did not show a statistically significant preference for a particular reading direction in the horizontal direction as predicted by the physical world account, descriptively more British children (57.7%) counted from left to right than right to left. I argue that this might be a hint of the emergence of the effect of reading direction on horizontal counting direction in British children.

## EXPERIMENT 2

Experiment 2 tested the flexibility of the counting pattern. Previous research (Bächtold et al., 1998; Ristic et al., 2006; Fischer et al., 2009, 2010) has shown that the SNARC effect is flexible and can be altered easily by short spatial experiences. For example, in a study by Shaki and Fischer (2008) bilingual Russian-Hebrew speakers showed a significant horizontal SNARC effect after reading a Russian text for 10 min (written in Cyrillic, reading direction left-to-right), but a significantly smaller horizontal SNARC effect after reading a Hebrew text (reading direction right-to-left) for the same amount of time. Inspired by this study, we asked HK-Chinese students living in the UK to count objects arranged in a 6×6 grid after they read a horizontal or vertical text. The reading direction account predicts that overall, in line with the dominant reading direction, the majority of participants will count the objects from top left to bottom right, row by row. In addition, it is expected that more participants will count from top right to bottom left, column by column, after reading the vertical text than after reading the horizontal text. A second aim of the study was to investigate whether, similarly to Dehaene et al. (1993), the length of stay in the UK also influenced the strength of the vertical SNA.

## MATERIALS AND METHODS

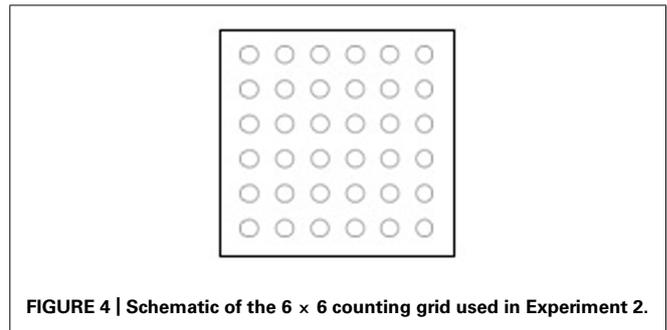
### Participants

Ninety-three right-handed native Cantonese speakers (18–25 years old, mean age = 20.63, SD = 1.27, 57 female, 36 male), brought up in HK, were tested. All had been living in the UK for less than 5 years (between 1 month and 5 years, mean years = 2.80, median = 3.00, SD = 1.43) and had given written consent. The study was approved by the Ethics Committee, Department of Psychology, University of York.

### Materials

The display consisted of 36 identical black unfilled circles (circumference = 1.6 cm) on a white piece of paper (19.2 cm by 19.2 cm). Circles were presented in a 6 by 6 grid with each circle at approximately 1.6 cm from the next circle and the outer circles at 0.4 cm from the edge (see **Figure 4**).

The reading material was a one-page text on attitudes about facing loss in Cantonese. It was taken from a website for Chinese reading comprehension (MaMa Resources, 2012; Supplementary material A, B, and C). Six comprehension questions were presented



**FIGURE 4 |** Schematic of the 6 × 6 counting grid used in Experiment 2.

on a separate sheet of paper. There were two conditions: a vertical and a horizontal text condition. In the vertical text condition, text on all three pages (the consent form, the short article and the comprehension questions) was presented in a vertical layout. For this text presentation the reader starts at the top right corner, reading column by column top to bottom, moving from right to left for each subsequent column. In the horizontal text condition all text was presented in horizontal layout that could only be read by starting at the top left corner moving from left to right in each row, starting with the top row and reading downward row by row from the top to the bottom row. The content of the horizontally and vertically presented consent form, article, and comprehension questions was identical (see Supplementary material A and B).

### Procedure

Participants were tested individually in the UK. Upon arrival participants were pseudorandomly allocated to either the vertical or horizontal reading condition and were tested individually in a quiet room in Cantonese by a native Cantonese speaker. Participants were asked to take a seat at the table where the stimuli had already been placed and covered before the participant arrived. Then, they were given the text to read. Subsequently they were given a sheet with comprehension questions and a pen and asked to provide the answers to the questions in writing. Participants in the vertical reading condition were given the consent form, text and comprehension questions in vertical layout, while participants in the horizontal reading condition received the consent form, text, and comprehension questions in horizontal layout. After the reading task, all participants were presented with the counting task. Participants were asked to count aloud the number of circles present on the piece of paper as quickly as possible while pointing to each circle. It was emphasized that even if it was obvious how many dots the display contained, they should still point to and count each circle. No demonstration was given, and the participants' order of counting was recorded by the experimenter. The seating position of the experimenter (to the left or right of the participant) was counterbalanced between participants.

## RESULTS

Six participants were excluded from the data analysis because their starting position for counting was not at the top, bottom, left, or right side of the grid.

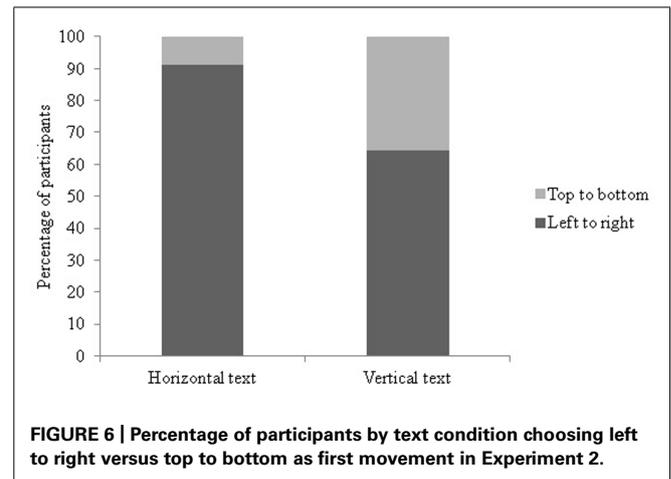
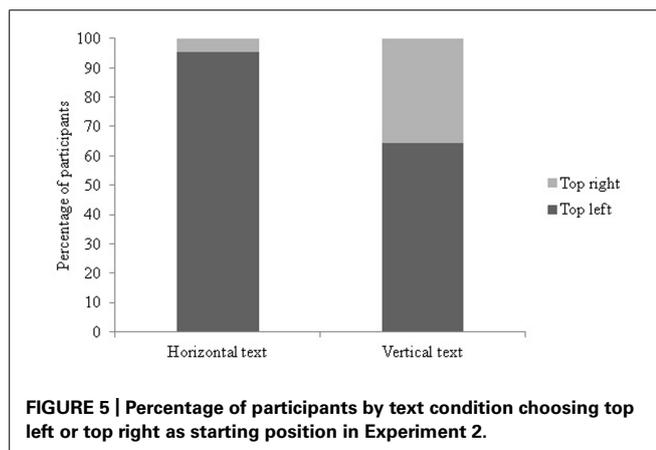
### Starting position

All remaining participants started counting at a top position. Most participants started counting at the top left of the grid. 70 participants (80.5%) started counting on the top left and 17 participants (19.5%) started on the top right side of the grid ( $\chi^2 = 32.87$ ,  $df = 1$ ,  $p < 0.01$ ). We then split participants into two groups depending on the length of stay in the UK (median split: short: less than 3 years; longer: 3 years or longer). In line with our predictions the length of time spent in the UK had a significant effect on their starting position ( $\chi^2 = 4.41$ ,  $df = 1$ ,  $p < 0.05$ , see Supplementary Table S2): although in both groups the most frequent starting position was top left, there were significantly more participants in the short stay (31.3%) than in the longer stay group (12.7%) starting counting at the top right.

In line with our predictions, there was a significant effect of text direction on the starting position ( $\chi^2 = 13.50$ ,  $df = 1$ ,  $p < 0.01$ , see Figure 5): in the horizontal text group 95.6% of participants started counting on the top left while in the vertical text condition it was only 64.3%. For the horizontal group there was also a significant effect of length of stay in the UK on the starting position ( $\chi^2 = 5.76$ ,  $df = 1$ ,  $p < 0.05$ ): none of the participants who had been in the UK three years or longer started counting on the top right, while 16.7% of the participants who arrived within the last three years did. There was no significant effect of length of stay on starting position for the vertical text group ( $\chi^2 = 0.31$ ,  $df = 1$ ,  $p = 0.58$ ).

### First movement

As expected, the majority of participants started counting with a horizontal movement. Sixty-eight participants (78.2%) moved horizontally from their starting position, 19 (21.8%) vertically ( $\chi^2 = 27.60$ ,  $df = 1$ ,  $p < 0.01$ , see Supplementary Table S2). All of the participants moving horizontally moved from left to right, and all of the participants moving vertically moved from top to bottom. In line with our predictions there was a significant effect of text direction on the direction of the first movement ( $\chi^2 = 9.16$ ,  $df = 1$ ,  $p < 0.01$ , see Figure 6): in the vertical text group 35.7% of participants started counting top to bottom while in the horizontal condition it was only 8.9%. Length of stay in the UK had no significant effect on the direction of the first movement ( $\chi^2 = 2.63$ ,  $df = 1$ ,  $p = 0.11$ ).



### Gender and experimenter seating position

There were no significant effects of gender and seating position of the experimenter on starting position or first movement (all  $p > 0.09$ ).

### DISCUSSION

Experiment 2 showed that directional reading habits dominate the counting behavior of adults. In line with their dominant reading direction, most HK-Chinese adults counted a square of circles from top left to bottom right, row by row. However, the frequency of this counting pattern was modulated by two factors: first, the most recent reading experience and second, reading experience within the last few years. Although most participants who had just read a vertical text still showed a preference for counting from top left to bottom right and row by row, significant more participants counted from top right to bottom left, column by column, after reading a vertical text than after reading the horizontal text. This is direct evidence for an influence of the most recent reading experience on the pattern of counting. Secondly, there is some evidence that the strength of this effect can be influenced by how long participants had lived in the UK: while nobody who had stayed in the UK longer than 3 years counted vertically after reading the horizontal text, two participants who had stayed in the UK less than 3 years at the time of testing did so.

### GENERAL DISCUSSION

In summary, our results strongly support the reading direction account. The majority of British and HK-Chinese participants counted the horizontal array in line with their dominant reading direction, from left to right. The vertical array in turn, they mostly counted from top to bottom, in line with their secondary reading direction, highlighting that both, horizontal and vertical, aspects of reading influence the direction of counting objects. Finally, when counting the four coins arranged in a square, the majority started on the top left coin, moved from the left to the right coin and then to the bottom left coin before ending on the bottom right coin. This pattern parallels a typically pattern of reading (Western) text on a page. Even the divergent results from the UK children fit with the reading direction account: I argue that UK children did not show a significant preference for a particular horizontal

counting direction yet, because the influence of reading direction is still weak at that age due to their limited reading and writing experience. Similarly, UK children still displayed a bottom-to-top counting preference for the vertical array. I suggest that they showed this pattern because their experience with and thus the influence of the secondary reading direction is even weaker and in its absence the experience of magnitude in the physical world dominates the vertical SNA. While Experiment 1 investigated the effect of reading direction on counting by comparing groups of participant with different reading experiences, in Experiment 2 we directly manipulated the frequency of counting direction by varying the most recent reading direction. Significantly more participants counted from top right moving top to bottom, column by column, after reading a vertical text than after reading a horizontal text. These results provide direct experimental evidence of an effect of the most recent reading direction on the direction of counting. Overall, our results are best explained by the reading direction account.

However, there are alternative accounts of the origins of horizontal and vertical SNAs which will be examined in the following sections. Although horizontal SNAs seem to be weaker in younger children, several recent studies have reported the existence of horizontal SNAs in infants (de Hevia et al., 2014) and animals (Adachi, 2014; Drucker and Brannon, 2014; Rugani et al., 2014, 2015). These findings are difficult to reconcile with the reading direction account and support a biological rather than cultural account, at least for early horizontal SNAs. The hemispheric lateralization account, however, cannot explain why in our study UK children show the horizontal SNA less strongly than HK-Chinese children unless one postulates that hemispheric lateralization is stronger in HK-Chinese children than in UK children of the same age which seems unlikely. Second, the hemispheric lateralization account predicts a left-to-right counting bias in both literate and illiterate adults and does not account for the reversal in participants who read from right to left. Clearly, the hemispheric lateralization account on its own is insufficient to explain the existing data on cultural counting direction.

However, a recent study (de Hevia et al., 2014) provides a suggestion for how the hemispheric lateralization account and the reading direction account of SNAs could be reconciled (a *combination account*). They found a preference for numerical increasing sequences from left to right in 7-month-old infants. This preference was context-dependent: it was only present when infants received the increasing condition before the decreasing condition, but not when the presentation order was reversed. This suggests that there might be a biological predisposition to link numerical order to spatial directionality *and* that this early bias is easily modifiable by experiential and cultural factors such as reading direction (de Hevia et al., 2012; for an overview of other early experiential and cultural factors see Nuerk et al., 2015).

Another factor for the origin of horizontal SNAs has been suggested by Fischer and Brugger (2011): finger counting habits. Fischer (2012, p. 163) cites finger counting habits and its relationship with horizontal SNAs as an example of embodiment, “sensory and/or motor constraints of the human body,” shaping number concepts. In an online survey of over 900 adults

(Lindemann et al., 2011) the majority of Western participants reported starting counting with their left hand while the majority of Eastern participants started with their right hand. These finger counting habits are in line with the direction of their horizontal SNAs. However, this study cannot discern between two options: finger counting habits could shape the direction of horizontal SNAs or vice versa. The crucial test is whether children’s finger counting direction is predictive of their dominant object counting direction. Recent findings by (Knudsen et al., in press) suggest that the answer is likely to be ‘no’: the majority of German 6-year-old children tested started counting with fingers on their right hand, but displayed a significant preference to count objects from left to right. In addition, finger counting habits cannot explain horizontal SNAs in animals and preverbal infants.

A clear advantage of the reading direction account is that it can explain SNAs in both horizontal and vertical dimensions. Both, the finger counting habits account and the hemispheric lateralization account, cannot explain counting preferences in the vertical dimension. The preferred vertical counting direction of adults and HK-Chinese children in our study is in line with their secondary reading direction and opposite to the direction predicted by the physical world account. This is puzzling, because most studies of the vertical SNARC effect have reported a bottom-to-top orientation for increasing magnitude (Gevers et al., 2006; Müller and Schwarz, 2007; Holmes and Lourenco, 2012; Shaki and Fischer, 2012; Hartmann et al., 2014). Why did we find a clear top-to-bottom association in the vertical array for our adult participants and HK-Chinese children when most vertical SNAs have been reported to go from bottom to top? Why should a reading direction account explain vertical counting direction while the physical world account is used to explain the vertical SNARC effect? I argue that these divergent results are due to two reasons: first, in contrast to parity judgment in the SNARC experiments, counting objects is in itself a spatial and explicitly numerical activity, so with object counting we are testing explicit associations between number and space which might be different from implicit associations between number and space tested in the SNARC effect (see Nuerk et al., 2015). Second, I propose that the required spatial movement inherent in counting objects in the sagittal plane activates reading experience more strongly than choosing one of two spatial response buttons in a parity judgment task. At least in initial stages of reading, people often use their fingers to guide them when reading text on a page. Similarly, when counting objects in space, participants used their fingers to point to objects in space. I argue that object counting *per se* is a spatial activity that automatically activates magnitude and that particularly in the horizontal plane we used, at least in competent readers, this space is strongly associated with reading and writing.

In our study the group with the smallest reading and writing experience, UK children, preferred to count the vertical array from bottom to top. This association of bottom with small and top with larger magnitude can neatly be explained by the physical world account: in our daily interactions with the physical world there are many examples of experiences where ‘more is up’ (Lakoff and Núñez, 2000; Hartmann et al., 2014) with the ground

level providing a natural zero (Holmes and Lourenco, 2012). Fischer (2012) argues that this is an example of grounded cognition (Barsalou, 2008). A higher mountain takes more effort and more time to climb than a smaller one. In contrast to our data, the physical world account does not predict cultural differences in vertical SNAs because the experience of the physical world is universal: the same physical principles apply independent of geographical location on our planet.

Perhaps related to experiences of magnitude in the physical world (Barsalou, 2008; Lachmair et al., 2014), we commonly encounter and use linguistic metaphors (Pecher and Boot, 2011) that associate more with higher, for example, 'prices rise' and 'I'll just turn up the volume.' These linguistic factors have spatial consequences. After reading descriptions of magnitudes (more or less) in sentences participants were faster to respond with a top button after 'more' sentences and a bottom button after 'less' sentences (Sell and Kaschak, 2012). Even in an unrelated categorization task after judging magnitudes (few or many?) participants responded faster after a 'many' judgment when the item to be categorized was presented at the top of the screen than when it was presented at the bottom (Pecher and Boot, 2011). The current study does not allow us to distinguish between the physical world account and the linguistic metaphor account for vertical SNAs in inexperienced readers, because both accounts predict an association of smaller magnitude with bottom space and larger magnitude with top space. However, a study by Holmes and Lourenco (2012) indicates that the vertical direction might be less malleable by verbal (metaphorical) instruction than has been reported for the horizontal direction (e.g., Bächtold et al., 1998; Ristic et al., 2006; Fischer et al., 2010). During a parity judgment task they asked participants explicitly to think of numbers as floors of a building (bottom-to-top metaphor), as items on a shopping list (top-to-bottom metaphor) or as diving levels in a swimming pool (top-to-bottom metaphor). In all three conditions participants associated smaller numbers with bottom and larger numbers with top space. Following a physical world account for vertical SNAs, one might expect the association of small magnitude with bottom and large magnitude with top space to be strong, stable, fixed, and unaltered by instruction, because the universal physical principles on our planet (e.g., gravity) almost never change. However, our findings in Experiment 2 suggest that the vertical counting direction can be modified by recent reading direction. Also, the vertical SNARC effect can be modified by effector instruction (Müller and Schwarz, 2007) and by different number notations (Hung et al., 2008). These results speak against a fixed vertical SNA with a grounded origin and provide good evidence that vertical SNAs can also be altered by instruction and recent experiences (see Hartmann et al., 2014).

This takes us to the question of whether counting direction preferences in a truly vertical plane would be different. Most studies on vertical SNAs, including the current study, have not used a truly vertical plane but a horizontal plane with close and far locations. The horizontal plane is heavily used when reading and writing, thus favoring a situated conception. A truly vertical plane might be a better test of the physical world account for vertical associations. Two recent studies have used a truly vertical plane

(Holmes and Lourenco, 2012; Hartmann et al., 2014) and reported a bottom-to-top association. To our knowledge, counting in the truly vertical direction, e.g., counting a stack of blocks has not been investigated systematically yet.

In both experiments presented here participants were asked to point to the objects and count them. On the basis of the current experiments it is not possible to exclude the possibility that pointing alone (without counting) could have resulted in spatial preferences too. Non-numerical horizontal spatial directional training can lead to changes in directional motor behavior in a visual search task (Patro et al., in press). Furthermore, culture-dependent biases in line bisection (Chokron and De Agostini, 1995; Rinaldi et al., 2014) as well as a culture-dependent preferences for the direction of drawing (Kebbe and Vinter, 2013) have been reported for the horizontal direction. So it is plausible that culture-dependent preferences in performing motor actions (such as pointing) might have contributed to the counting bias. Future studies should investigate directional preferences for both counting and pointing.

In summary, I have discussed the evidence for grounded, embodied, and situated origins of horizontal and vertical SNAs. A *combination account* (de Hevia et al., 2012; Nuerk et al., 2015) is emerging: due to hemispheric lateralization (Rugani et al., 2014, 2015) and a preference for increasing magnitudes (Macchi Cassia et al., 2012) we start life with a slight preference to associate small magnitudes with the left side of space (a biological predisposition). In addition, interactions with the physical world (grounded cognition; Barsalou, 2008) lead us to expect magnitudes to increase from the bottom to the top resulting in an initial SNA with increasing magnitude from bottom left to top right. Interactions with cultural spatial biases in the environment such as exposure to cultural reading practices then modify this initial bias: depending on the culturally predominant spatial directionality the bottom left to top right bias either gets strengthened, weakened, or overwritten. Although further research into other cultural spatial biases is needed, current evidence favors reading direction as the strongest cultural spatial influence. SNAs molded by longstanding cultural directional biases can also be modified temporarily by recent spatial experiences.

To conclude, our findings clearly support the influence of primary and secondary reading direction on the horizontal and vertical direction of counting in the horizontal plane and its relationship to recent as well as longstanding reading exposure and experience.

## ACKNOWLEDGMENTS

I would like to thank the children, adults, nurseries, and schools who took part in this study as well as Allison Cheung, Kelsey Lam, Carmen Lau, Jasmine Lei, Rachel Meyrick, Courtney Poole, and Ellen Teder for their assistance in collecting and coding the data. SG was supported by a British Academy/Leverhulme Small Research Grant (SG121544).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fpsyg.2015.00228/abstract>

## REFERENCES

- Adachi, I. (2014). Spontaneous spatial mapping of learned sequence in chimpanzees: evidence for a SNARC-like effect. *PLoS ONE* 9:e90373. doi: 10.1371/journal.pone.0090373
- Bächtold, D., Baumüller, M., and Brugger, P. (1998). Stimulus–response compatibility in representational space. *Neuropsychologia* 36, 731–735. doi: 10.1016/S0028-3932(98)00002-5
- Barsalou, L. W. (2008). Grounded cognition. *Annu. Rev. Psychol.* 59, 617–645. doi: 10.1146/annurev.psych.59.103006.093639
- Berch, D. B., Foley, E. J., Hill, R. J., and Ryan, P. M. (1999). Extracting parity and magnitude from Arabic numerals: developmental changes in number processing and mental representation. *J. Exp. Child Psychol.* 74, 286–308. doi: 10.1006/jecp.1999.2518
- Chokron, S., and De Agostini, M. (1995). Reading habits and line bisection: a developmental approach. *Cogn. Brain Res.* 3, 51–58. doi: 10.1016/0926-6410(95)00018-6
- Curriculum Development Council. (2006). *Guide to the Pre-primary Curriculum*. Hong Kong: Curriculum Development Council.
- Dehaene, S., Bossini, S., and Giraux, P. (1993). The mental representation of parity and numerical magnitude. *J. Exp. Psychol. Gen.* 122, 371–396. doi: 10.1037/0096-3445.122.3.371
- de Hevia, M. D., Girelli, L., Addabbo, M., and Macchi Cassia, V. (2014). Human infants' preference for left-to-right oriented increasing numerical sequences. *PLoS ONE* 9:e96412. doi: 10.1371/journal.pone.0096412
- de Hevia, M. D., Girelli, L., and Macchi Cassia, V. (2012). Minds without language represent number through space: origins of the mental number line. *Front. Psychol.* 3:466. doi: 10.3389/fpsyg.2012.00466
- Dowker, A., Bala, S., and Lloyd, D. (2008). Linguistic influences on mathematical development: how important is the transparency of the counting system? *Philos. Psychol.* 21, 523–538. doi: 10.1080/09515080802285511
- Drucker, C. B., and Brannon, E. M. (2014). Rhesus monkeys (*Macaca mulatta*) map number onto space. *Cognition* 132, 57–67. doi: 10.1016/j.cognition.2014.03.011
- Fischer, M. H. (2012). A hierarchical view of grounded, embodied, and situated numerical cognition. *Cogn. Process.* 13, 161–164. doi: 10.1007/s10339-012-0477-5
- Fischer, M. H., and Brugger, P. (2011). When digits help digits: spatial–numerical associations point to finger counting as prime example of embodied cognition. *Front. Psychol.* 2:260. doi: 10.3389/fpsyg.2011.00260
- Fischer, M. H., Mills, R. A., and Shaki, S. (2010). How to cook a SNARC: number placement in text rapidly changes spatial–numerical associations. *Brain Cogn.* 72, 333–336. doi: 10.1016/j.bandc.2009.10.010
- Fischer, M. H., Shaki, S., and Cruise, A. (2009). It takes just one word to quash a SNARC. *Exp. Psychol.* 56, 361–366. doi: 10.1027/1618-3169.56.5.361
- Galton, F. (1880). Visualised numerals. *Nature* 22, 252–256. doi: 10.1038/021252a0
- Gevers, W., Lammertyn, J., Notebaert, W., Verguts, T., and Fias, W. (2006). Automatic response activation of implicit spatial information: evidence from the SNARC effect. *Acta Psychol.* 122, 221–233. doi: 10.1016/j.actpsy.2005.11.004
- Göbel, S. M., Calabria, M., Farnè, A., and Rossetti, Y. (2006). Parietal rTMS distorts the mental number line: simulating 'spatial' neglect in healthy subjects. *Neuropsychologia* 44, 860–868. doi: 10.1016/j.neuropsychologia.2005.09.007
- Göbel, S. M., Fischer, M. H., and Shaki, S. (2014a). "Experience counts: transmitting culture through space," in *Poster and Oral Presentation at the 32nd European Workshop on Cognitive Neuropsychology*, Bressanone.
- Göbel, S. M., Moeller, K., Pixner, S., Kaufmann, L., and Nuerk, H. C. (2014b). Language affects symbolic arithmetic in children: the case of number word inversion. *J. Exp. Child Psychol.* 119, 17–25. doi: 10.1016/j.jecp.2013.10.001
- Göbel, S. M., Shaki, S., and Fischer, M. H. (2011). The cultural number line: a review of cultural and linguistic influences on the development of number processing. *J. Cross Cult. Psychol.* 42, 543–565. doi: 10.1177/0022022111406251
- Hartmann, M., Gashaj, V., Stahnke, A., and Mast, F. W. (2014). There is more than "more is up": hand and foot responses reverse the vertical association of number magnitudes. *J. Exp. Psychol. Hum. Percept. Perform.* 40, 1401–1414. doi: 10.1037/a0036686
- Heilman, K. M., and Van Den Abell, T. (1980). Right hemisphere dominance for attention: the mechanism underlying hemispheric asymmetries of inattention (neglect). *Neurology* 30, 327–327. doi: 10.1212/WNL.30.3.327
- Hoffmann, D., Hornung, C., Martin, R., and Schiltz, C. (2013). Developing number–space associations: SNARC effects using a color discrimination task in 5-year-olds. *J. Exp. Child Psychol.* 116, 775–791. doi: 10.1016/j.jecp.2013.07.013
- Holmes, K. J., and Lourenco, S. F. (2012). Orienting numbers in mental space: horizontal organization trumps vertical. *Q. J. Exp. Psychol.* 65, 1044–1051. doi: 10.1080/17470218.2012.685079
- Hung, Y.-H., Hung, D. L., Tzeng, O. J.-L., and Wu, D. H. (2008). Flexible spatial mapping of different notations of numbers in Chinese readers. *Cognition* 106, 1441–1450. doi: 10.1016/j.cognition.2007.04.017
- Ito, Y., and Hatta, T. (2004). Spatial structure of quantitative representation of numbers: evidence from the SNARC effect. *Mem. Cogn.* 32, 662–673. doi: 10.3758/BF03195857
- Kebbe, H., and Vinter, A. (2013). How culture, age, and manual dominance affect directionality in drawing side view objects. *J. Cross Cult. Psychol.* 44, 160–172. doi: 10.1177/0022022111435098
- Knudsen, B., Fischer, M. H., and Aschersleben, G. (in press). Development of spatial preferences for counting and picture naming. *Psychol. Res.* 1–11. doi: 10.1007/s00426-014-0623-z
- Lachmair, M., Dudschig, C., de la Vega, I., and Kaup, B. (2014). Relating numeric cognition and language processing: do numbers and words share a common representational platform? *Acta Psychol.* 148, 107–114. doi: 10.1016/j.actpsy.2013.12.004
- Lakoff, G., and Núñez, R. (2000). *Where Mathematics Comes From*. New York: Basic Books.
- Lindemann, O., Alipour, A., and Fischer, M. H. (2011). Finger counting habits in middle eastern and western individuals: an online survey. *J. Cross Cult. Psychol.* 42, 566–578. doi: 10.1177/0022022111406254
- Macchi Cassia, V., Picozzi, M., Girelli, L., and de Hevia, M. D. (2012). Increasing magnitude counts more: asymmetrical processing of ordinality in 4-month-old infants. *Cognition* 124, 183–193. doi: 10.1016/j.cognition.2012.05.004
- MaMa Resources. (2012). 中一, 閱讀理解練習: 坦然面對. Available at: <http://mamaresort.blogspot.co.uk/2012/12/blog-post.html> (accessed December 17, 2012).
- Mesulam, M. (1990). Large scale neurocognitive networks and distributed processing for attention, language, and memory. *Ann. Neurol.* 28, 597–613. doi: 10.1002/ana.410280502
- Müller, D., and Schwarz, W. (2007). Is there an internal association of numbers to hands? The task set influences the nature of the SNARC effect. *Mem. Cogn.* 35, 1151–1161. doi: 10.3758/BF03193485
- Nuerk, H.-C., Patro, K., Cress, U., Schild, U., Friedrich, C. K., and Göbel, S. M. (2015). How space–number associations may be created in preliterate children: six distinct mechanisms. *Front. Psychol.* 6:215. doi: 10.3389/fpsyg.2015.00215
- Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9, 97–113. doi: 10.1016/0028-3932(71)90067-4
- Opfer, J. E., and Thompson, C. A. (2006). "Even early representations of numerical magnitude are spatially organized: evidence from a directional magnitude bias in pre-reading preschoolers," in *Proceedings of the 28th annual cognitive science society*, eds R. Sun and N. Miyake (Mahwah, NJ: Erlbaum), 639–644.
- Opfer, J. E., Thompson, C. A., and Furlong, E. E. (2010). Early development of spatial–numeric associations: evidence from spatial and quantitative performance of preschoolers. *Dev. Sci.* 13, 761–771. doi: 10.1111/j.1467-7687.2009.00934.x
- Patro, K., Fischer, U., Nuerk, H.-C., and Cress, U. (in press). How to rapidly construct a spatial–numerical representation in preliterate children (at least temporarily). *Dev. Sci.*
- Patro, K., and Haman, M. (2012). The spatial–numerical congruity effect in preschoolers. *J. Exp. Child Psychol.* 111, 534–542. doi: 10.1016/j.jecp.2011.09.006
- Pecher, D., and Boot, I. (2011). Numbers in space: differences between concrete and abstract situations. *Front. Psychol.* 2:121. doi: 10.3389/fpsyg.2011.00121
- Pixner, S., Moeller, K., Hermanova, V., Nuerk, H. C., and Kaufmann, L. (2011). Whorf reloaded: language effects on nonverbal number processing in first grade—A trilingual study. *J. Exp. Child Psychol.* 108, 371–382. doi: 10.1016/j.jecp.2010.09.002
- Rinaldi, L., Di Luca, S., Henik, A., and Girelli, L. (2014). Reading direction shifts visuospatial attention: an interactive account of attentional biases. *Acta Psychol.* 151, 98–105. doi: 10.1016/j.actpsy.2014.05.018
- Ristic, J., Wright, A., and Kingstone, A. (2006). The number line effect reflects top-down control. *Psychon. Bull. Rev.* 13, 862–868. doi: 10.3758/BF03194010
- Rugani, R., Kelly, D. M., Szelest, I., Regolin, L., and Vallortigara, G. (2010). Is it only humans that count from left to right? *Biol. Lett.* 6, 290–292. doi: 10.1098/rsbl.2009.0960

- Rugani, R., Salva, O. R., and Regolin, L. (2014). Lateralized mechanisms for encoding of object. Behavioral evidence from an animal model: the domestic chick (*Gallus gallus*). *Front. Psychol.* 5:150. doi: 10.3389/fpsyg.2014.00150
- Rugani, R., Vallortigara, G., Pifrits, K., and Regolin, L. (2015). Number-space mapping in the newborn chick resembles humans' mental number line. *Science* 347, 534–536. doi: 10.1126/science.aaa1379
- Rugani, R., Vallortigara, G., Vallini, B., and Regolin, L. (2011). Asymmetrical number-space mapping in the avian brain. *Neurobiol. Learn. Mem.* 95, 231–238. doi: 10.1016/j.nlm.2010.11.012
- Sagiv, N., Simner, J., Collins, J., Butterworth, B., and Ward, J. (2006). What is the relationship between synaesthesia and visuo-spatial number forms? *Cognition* 101, 114–128. doi: 10.1016/j.cognition.2005.09.004
- Schwarz, W., and Keus, I. M. (2004). Moving the eyes along the mental number line: comparing SNARC effects with saccadic and manual responses. *Percept. Psychophys.* 66, 651–664. doi: 10.3758/BF03194909
- Sell, A. J., and Kaschak, M. P. (2012). The comprehension of sentences involving quantity information affects responses on the up–down axis. *Psychon. Bull. Rev.* 19, 708–714. doi: 10.3758/s13423-012-0263-5
- Seron, X., Pesenti, M., Noël, M. P., Deloche, G., and Cornet, J. A. (1992). Images of numbers, or “When 98 is upper left and 6 sky blue.” *Cognition* 44, 159–196. doi: 10.1016/0010-0277(92)90053-K
- Shaki, S., and Fischer, M. H. (2008). Reading space into numbers—a cross-linguistic comparison of the SNARC effect. *Cognition* 108, 590–599. doi: 10.1016/j.cognition.2008.04.001
- Shaki, S., and Fischer, M. H. (2012). Multiple spatial mappings in numerical cognition. *J. Exp. Psychol.* 38, 804–809. doi: 10.1037/a002756
- Shaki, S., Fischer, M. H., and Göbel, S. M. (2012). The origin of number-space associations: a comparative study of spatially directional counting biases in cultures with different reading directions. *J. Exp. Child Psychol.* 112, 275–281. doi: 10.1016/j.jecp.2011.12.005
- Shaki, S., Fischer, M. H., and Petrusic, W. M. (2009). Reading habits for both words and numbers contribute to the SNARC effect. *Psychon. Bull. Rev.* 16, 328–331. doi: 10.3758/PBR.16.2.328
- Siegler, R. S., and Mu, Y. (2008). Chinese children excel on novel mathematics problems even before elementary school. *Psychol. Sci.* 19, 759–763. doi: 10.1111/j.1467-9280.2008.02153.x
- Umiltà, C., Pifrits, K., and Zorzi, M. (2009). The spatial representation of numbers: evidence from neglect and pseudoneglect. *Exp. Brain Res.* 192, 561–569. doi: 10.1007/s00221-008-1623-2
- Van Galen, M. S., and Reitsma, P. (2008). Developing access to number magnitude: a study of the SNARC effect in 7- to 9-year-olds. *J. Exp. Child Psychol.* 101, 99–113. doi: 10.1016/j.jecp.2008.05.001
- Wood, G., Willmes, K., Nuerk, H. C., and Fischer, M. H. (2008). On the cognitive link between space and number: a meta-analysis of the SNARC effect. *Psychol. Sci. Q.* 50, 489–525.
- Zebian, S. (2005). Linkages between number concepts, spatial thinking, and directionality of writing: the SNARC effect and the reverse SNARC effect in English and Arabic monoliterates, biliterates, and illiterate Arabic speakers. *J. Cogn. Cult.* 5, 165–190. doi: 10.1163/1568537054068660
- Zuber, J., Pixner, S., Moeller, K., and 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. *J. Exp. Child Psychol.* 102, 60–77. doi: 10.1016/j.jecp.2008.04.003

**Conflict of Interest Statement:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 27 November 2014; accepted: 14 February 2015; published online: 10 March 2015.

Citation: Göbel SM (2015) Up or down? Reading direction influences vertical counting direction in the horizontal plane – a cross-cultural comparison. *Front. Psychol.* 6:228. doi: 10.3389/fpsyg.2015.00228

This article was submitted to *Developmental Psychology*, a section of the journal *Frontiers in Psychology*.

Copyright © 2015 Göbel. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Two steps to space for numbers

Martin H. Fischer<sup>1\*</sup> and Samuel Shaki<sup>2</sup>

<sup>1</sup> Division of Cognitive Sciences, Department of Psychology, University of Potsdam, Potsdam, Germany, <sup>2</sup> Department of Behavioral Science, Ariel University, Ariel, Israel

**Keywords:** spatial-numerical association, SNARC, mental number line, numerical cognition, spatial cognition

The study of spatial-numerical associations (SNAs) is an active research project that was triggered by a landmark publication reporting several simple reaction time experiments: Adults classified visually presented numbers according to their parity by using left and right response keys (Dehaene et al., 1993). The main finding was that small numbers, such as 1 or 2, were classified faster on the left side and larger numbers, such as 8 or 9, were classified faster on the right side. This specific instance of a SNA has been replicated and extended in numerous studies (recent review by Fischer and Shaki, 2014). The original interpretation of the effect assumed a “spill-over” from reading habits into the number domain but subsequent work has pushed back the time line to preschoolers, infants, and even neonates (for recent review, see Patro et al., 2014). Our own work (e.g., Shaki et al., 2009; Fischer and Shaki, 2015) confirmed that reading habits contribute to the direction and strength of SNAs but has also indicated that they are not the only and not even the strongest determinant (e.g., Fischer et al., 2010). In the following paragraphs we propose a processing principle for SNAs and describe two successive steps by which the mapping of numbers onto space might occur.

Our proposed processing principle is that *spatial mapping is an integral part of semantic number processing*. This is evident from the ubiquity of SNAs: They have been reported with various stimulus formats, in many different tasks, and while studying a wide range of responses (for recent review, see Fischer and Shaki, 2014). SNAs modulate the cortical region underlying semantic number processing (i.e., bilateral HIPS; Cutini et al., 2012). Moreover, the association between numbers and space is bi-directional: numerical magnitude can serve as a spatial cue and vice versa (Stoianov et al., 2008; Shaki and Fischer, 2014a). Most studies of SNAs have used centrally presented numbers in combination with spatial responses, which may have encouraged participants to use spatial number mapping strategies (Fischer, 2006). However, today it is clear that the very appearance of numerical stimuli is enough for SNAs to appear, even when removing, in healthy adults (cf. Zorzi et al., 2002), spatial features from *both* stimuli and responses (Fischer and Shaki, 2015; Ranzini et al., 2015). Evidence for such a purely conceptual link between numbers and space was even found in Hebrew speakers, thus requiring correction of our earlier claim of the need for consistency of directional processing habits across stimulus domains (Shaki et al., 2009; Shaki and Fischer, 2012, 2014b).

We note that our present proposal leaves open the issue of the origin(s) of SNAs, be they a congenital result of hemispheric specializations, or acquired by culturally shaped spatial habits such as reading or finger counting (Fischer, 2008; Lindemann et al., 2011; Domahs et al., 2012; Fischer and Shaki, 2015; Rugani et al., 2015a,b). Assuming that processing number meaning is obligatorily accompanied by mapping it onto a spatial continuum, two issues remain to be addressed to account for a given SNA in a particular setting: The selection of the appropriate spatial dimension, and the directionality of mapping numbers along that dimension. We now present an idea of how these two steps are taken and describe recent evidence in support of this proposal.

First, *the spatial dimension selected for mapping of numbers reflects the stimulus and response features of the current task*. When lateralized response keys are provided to participants to measure the speed of their judgments, then most participants will align their number representations along the dimension indicated by these keys, be it horizontal, vertical, or radial. This is what the bulk of

## OPEN ACCESS

### Edited by:

Ann Dowker,  
University of Oxford, UK

### Reviewed by:

Marco Zorzi,  
University of Padova, Italy  
Katarzyna Patro,  
University of Tübingen, Germany and  
University of Warsaw, Poland

### \*Correspondence:

Martin H. Fischer,  
martinf@uni-potsdam.de

### Specialty section:

This article was submitted to  
Developmental Psychology,  
a section of the journal  
Frontiers in Psychology

**Received:** 02 November 2014

**Accepted:** 24 April 2015

**Published:** 12 May 2015

### Citation:

Fischer MH and Shaki S (2015) Two  
steps to space for numbers.  
Front. Psychol. 6:612.  
doi: 10.3389/fpsyg.2015.00612

the literature has documented (as recently reviewed by Fischer and Shaki, 2014). In the absence of such response keys, when responses to numbers are required by making spatially directional arm, head, eye or whole-body movements, then the major directions or endpoints of those movements define the mapping dimension, again either using the horizontal (Fischer, 2003; Fischer et al., 2004; Loetscher et al., 2008; Shaki and Fischer, 2014a) or vertical dimension (Schwarz and Keus, 2004; Winter and Matlock, 2013). When spatially distinct responses to the numbers are required but no response dimension is prescribed, the resulting mapping of numbers onto space will be more varied across participants (Fischer and Campens, 2008). Finally, even when no spatially distinct responses are required, as for example in a simple detection task, the spatial mapping of centrally presented numbers will still emerge through lateralization of other stimuli, such as visually presented cues (Fischer et al., 2003; for a recent update, see Fischer and Knops, 2014).

Finally, once a dimension for the spatial mapping of numbers has been selected by the participant, their distribution along this dimension still remains to be decided. For this second step, we propose that *the orientation of the SNA is influenced by spatial experience*. This rule underlines the manifold of possible influences on the SNARC which are only beginning to be documented and studied. Living in a three-dimensional world, we are differentially sensitive to horizontal vs. vertical space. For example, as a result of the embodied nature of cognition, vertical distinctions are most salient and horizontal ones least salient (Fischer and Brugger, 2011), leading to

faster acquisition of, and discrimination along, the vertical than the horizontal dimension (Franklin and Tversky, 1990). Similarly, the increasing strength of SNAs with age (Wood et al., 2008; Hoffmann et al., 2014) indicates that they may reflect accumulated spatial habits/experiences during life. An example are reading habits (see the contribution of Nuerk et al., 2015 to this research topic for a detailed description of mechanisms). Importantly, such life-long experiences are less powerful in determining the directionality of a SNAs compared to more recent experiences with numbers, as demonstrated in emerging training studies (e.g., Fischer, 2012) and by rapid alternations of SNAs between successive trials (Fischer et al., 2009).

In summary, the proposed two successive steps seem to capture a wide range of observations pertaining to the ubiquity of SNAs that have recently re-invigorated research into numerical cognition. We hope that the present proposal will guide further interest in the design of novel studies that aim to test specific predictions about the origin and strength of SNAs. For example, how can we identify the sequential nature of the mapping process? How shall we weight the contributions of previous experiences? Clearly, such questions identify numerical cognition as a convenient test-bed for the study of fundamental principles of cognition generally.

## Acknowledgment

MHF is funded by DFG grant FI 1915/2-1 “Manumerical cognition”.

## References

- Cutini, S., Scarpa, F., Scatturin, P., Dell'Acqua, R., and Zorzi, M. (2012). Number-space interactions in the human parietal cortex: enlightening the SNARC effect with functional near-infrared spectroscopy. *Cereb. Cortex* 24, 444–451. doi: 10.1093/cercor/bhs321
- Dehaene, S., Bossini, S., and Giraux, P. (1993). The mental representation of parity and number magnitude. *J. Exp. Psychol.* 122, 371–396. doi: 10.1037/0096-3445.122.3.371
- Domahs, F., Kaufmann, L., and Fischer, M. H. (2012). *Handy Numbers: Finger Counting and Numerical Cognition*. Lausanne: Special Research Topic in Frontiers Media SA.
- Fischer, M. H. (2003). Spatial representations in number processing—evidence from a pointing task. *Vis. Cogn.* 10, 493–508. doi: 10.1080/13506280244000186
- Fischer, M. H. (2006). The future for SNARC could be stark. *Cortex* 42, 1066–1068. doi: 10.1016/S0010-9452(08)70218-1
- Fischer, M.H. (2008). Finger counting habits modulate spatial-numerical associations. *Cortex* 44, 386–392. doi: 10.1016/j.cortex.2007.08.004
- Fischer, M. H. (2012). “The spatial mapping of numbers – its origin and flexibility,” in *Language and Action in Cognitive Neurosciences*, eds Y. Coello and A. Bartolo, (London: Psychology Press), 225–242.
- Fischer, M. H., and Brugger, P. (2011). When digits help digits: spatial-numerical associations point to finger counting as prime example of embodied cognition. *Front. Psychol.* 2:260. doi: 10.3389/fpsyg.2011.00260
- Fischer, M. H., and Campens, H. (2008). Pointing to numbers and grasping magnitudes. *Exp. Brain Res.* 192, 149–153. doi: 10.1007/s00221-008-1622-3
- Fischer, M. H., Castel, A. D., Dodd, M. D., and Pratt, J. (2003). Perceiving numbers causes spatial shifts of attention. *Nat. Neurosci.* 6, 555–556. doi: 10.1038/nn1066
- Fischer, M. H., and Knops, A. (2014). Attentional cueing in numerical cognition – a commentary on Zanolie and Pecher (2014). *Front. Psychol.* 5:1381. doi: 10.3389/fpsyg.2014.01381
- Fischer, M. H., Mills, R. A., and Shaki, S. (2010). How to cook a SNARC: number placement in text rapidly changes spatial-numerical associations. *Brain Cogn.* 72, 333–336. doi: 10.1016/j.bandc.2009.10.010
- Fischer, M. H., and Shaki, S. (2014). Spatial associations in numerical cognition: from single digits to arithmetic. *Q. J. Exp. Psychol.* 67, 1461–1483. doi: 10.1080/17470218.2014.927515
- Fischer, M. H., and Shaki, S. (2015). Measuring spatial-numerical associations: evidence for a purely conceptual link. *Psychol. Res.* doi: 10.1007/s00426-015-0646-0. [Epub ahead of print].
- Fischer, M. H., Shaki, S., and Cruise, A. (2009). It takes only one word to quash the SNARC. *Exp. Psychol.* 56, 361–366. doi: 10.1027/1618-3169.56.5.361
- Fischer, M. H., Warlop, N., Hill, R. L., and Fias, W. (2004). Oculomotor bias induced by number perception. *Exp. Psychol.* 51, 91–97. doi: 10.1027/1618-3169.51.2.91
- Franklin, N., and Tversky, B. (1990). Searching imagined environments. *J. Exp. Psychol.* 119, 63–76. doi: 10.1037/0096-3445.119.1.63
- Hoffmann, D., Pigat, D., and Schiltz, C. (2014). The impact of inhibition capacities and age on number-space associations. *Cogn. Process.* 15, 329–342. doi: 10.1007/s10339-014-0601-9
- Lindemann, O., Alipour, A., and Fischer, M. H. (2011). Finger counting habits in Middle-Eastern and Western individuals: an online survey. *J. Cross Cult. Psychol.* 42, 566–578. doi: 10.1177/0022022111406254
- Loetscher, T., Schwarz, U., Schubiger, M., and Brugger, P. (2008). Head turns bias the brain's internal random generator. *Curr. Biol.* 18, R60–R62. doi: 10.1016/j.cub.2007.11.015

- Nuerk, H.-C., Patro K., Cress, U., Schild, U., Friedrich, C. K., and Göbel, S. M. (2015). How space-number associations may be created in preliterate children: Six distinct mechanisms. *Front. Psychol.* 6:215. doi: 10.3389/fpsyg.2015.00215
- Patro, K., Nuerk, H.-C., Cress, U., and Haman, M. (2014). How number space relationships are assessed before formal schooling: a taxonomy proposal. *Front. Psychol.* 5:419. doi: 10.3389/fpsyg.2014.00419
- Ranzini, M., Lisi, M., Blini, E., Pitteri, M., Treccani, B., Priftis, K., et al. (2015). Larger, smaller, odd or even? Task-specific effects of optokinetic stimulation on the mental number space. *J. Cogn. Psychol.* 27, 459–470. doi: 10.1080/20445911.2014.941847
- Rugani, R., Vallortigara, G., Priftis, K., and Regolin, L. (2015a). Number-space mapping in the newborn chick resembles humans' mental number line. *Science* 347, 534–536. doi: 10.1126/science.aaa1379
- Rugani, R., Vallortigara, G., and Regolin, L. (2015b). At the root of the left-right asymmetries in spatial-numerical processing: from domestic chicks to human subjects. *J. Cogn. Psychol.* 27, 388–399. doi: 10.1080/20445911.2014.941846
- Schwarz, W., and Keus, I. (2004). Moving the eyes along the mental number line: comparing SNARC effects with manual and saccadic responses. *Percept. Psychophys.* 66, 651–664. doi: 10.3758/BF03194909
- Shaki, S., and Fischer, M. H. (2012). Multiple spatial mappings in numerical cognition. *J. Exp. Psychol. Hum. Percept. Perform.* 38, 804–809. doi: 10.1037/a0027562
- Shaki, S., and Fischer, M. H. (2014a). Random walks on the mental number line. *Exp. Brain Res.* 232, 43–49. doi: 10.1007/s00221-013-3718-7
- Shaki, S., and Fischer, M. H. (2014b). Removing spatial responses reveals spatial concepts – even in a culture with mixed reading habits. *Front. Hum. Neurosci.* 8:966. doi: 10.3389/fnhum.2014.00966
- Shaki, S., Fischer, M. H., and Petrusic, W. M. (2009). Reading habits for both words and numbers contribute to the SNARC effect. *Psychon. Bull. Rev.* 16, 328–331. doi: 10.3758/PBR.16.2.328
- Stoianov, I., Kramer, P., Umiltà, C., and Zorzi, M. (2008). Visuospatial priming of the mental number line. *Cognition* 106, 770–779. doi: 10.1016/j.cognition.2007.04.013
- Winter, B., and Matlock, T. (2013). “More is up... and right: random number generation along two axes,” in *Proceedings of the 35th Annual Conference of the Cognitive Science Society*, eds M. Knauff, M. Pauen, N. Sebanz, and I. Wachsmuth (Austin, TX: Cognitive Science Society), 3789–3974.
- Wood, G., Nuerk, H.-C., Willmes, K., and Fischer, M. H. (2008). On the cognitive link between space and number: a meta-analysis of the SNARC effect. *Psychol. Sci. Q.* 50, 489–525.
- Zorzi, M., Priftis, K., and Umiltà, C. (2002). Brain damage: neglect disrupts the mental number line. *Nature* 417, 138–139. doi: 10.1038/417138a

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Fischer and Shaki. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Spatial complexity of character-based writing systems and arithmetic in primary school: a longitudinal study

Maja Rodic<sup>1,2</sup>, Tatiana Tikhomirova<sup>2,3</sup>, Tatiana Koliienko<sup>2</sup>, Sergey Malykh<sup>2,4</sup>, Olga Bogdanova<sup>2</sup>, Dina Y. Zueva<sup>2</sup>, Elena I. Gynku<sup>2</sup>, Sirui Wan<sup>5</sup>, Xinlin Zhou<sup>5</sup> and Yulia Kovas<sup>1,2\*</sup>

<sup>1</sup> InLab, Department of Psychology, Goldsmiths, University of London, London, UK, <sup>2</sup> Laboratory for Cognitive Investigations and Behavioral Genetics, Department of Psychology, Tomsk State University, Tomsk, Russia, <sup>3</sup> Institute of Psychology, Russian Academy of Sciences, Moscow, Russia, <sup>4</sup> Psychological Institute, Russian Academy of Education, Moscow, Russia, <sup>5</sup> State Key Laboratory of Cognitive Neuroscience and Learning, Department of Psychology, Beijing Normal University, Beijing, China

## OPEN ACCESS

### Edited by:

Ann Dowker,  
University of Oxford, UK

### Reviewed by:

Phillip Dale,  
University of New Mexico, USA  
Chongying Wang,  
Nankai University, China

### \*Correspondence:

Yulia Kovas,  
InLab, Department of Psychology,  
Goldsmiths, University of London,  
SE14 6NW, Office WB312,  
London, UK  
y.kovas@gold.ac.uk

### Specialty section:

This article was submitted to  
Developmental Psychology, a section  
of the journal *Frontiers in Psychology*

**Received:** 30 November 2014

**Accepted:** 08 March 2015

**Published:** 26 March 2015

### Citation:

Rodic M, Tikhomirova T, Koliienko T,  
Malykh S, Bogdanova O, Zueva DY,  
Gynku EI, Wan S, Zhou X and Kovas  
Y (2015) Spatial complexity of  
character-based writing systems and  
arithmetic in primary school: a  
longitudinal study.  
*Front. Psychol.* 6:333.  
doi: 10.3389/fpsyg.2015.00333

Previous research has consistently found an association between spatial and mathematical abilities. We hypothesized that this link may partially explain the consistently observed advantage in mathematics demonstrated by East Asian children. Spatial complexity of the character-based writing systems may reflect or lead to a cognitive advantage relevant to mathematics. Seven hundred and twenty one 6–9-year old children from the UK and Russia were assessed on a battery of cognitive skills and arithmetic. The Russian children were recruited from specialist linguistic schools and divided into four different language groups, based on the second language they were learning (i.e., English, Spanish, Chinese, and Japanese). The UK children attended regular schools and were not learning any second language. The testing took place twice across the school year, once at the beginning, before the start of the second language acquisition, and once at the end of the year. The study had two aims: (1) to test whether spatial ability predicts mathematical ability in 7–9 year-old children across the samples; (2) to test whether acquisition and usage of a character-based writing system leads to an advantage in performance in arithmetic and related cognitive tasks. The longitudinal link from spatial ability to mathematics was found only in the Russian sample. The effect of second language acquisition on mathematics or other cognitive skills was negligible, although some effect of Chinese language on mathematical reasoning was suggested. Overall, the findings suggest that although spatial ability is related to mathematics at this age, one academic year of exposure to spatially complex writing systems is not enough to provide a mathematical advantage. Other educational and socio-cultural factors might play a greater role in explaining individual and cross-cultural differences in arithmetic at this age.

**Keywords:** early arithmetic, cross-cultural, longitudinal, character-based writing system, spatial ability

## Introduction

Research has shown that East Asian children on average outperform other children in mathematics (Miura, 1987; Song and Ginsburg, 1988; Stevenson and Stigler, 1992; Geary et al., 1993; Stevenson et al., 1993; Imbo and Vandierendonck, 2007; Mullis et al., 2008; OECD, 2010; Rodic et al., 2014). This advantage might partly be explained by the regular structure of the East Asian number system, as well as by the shorter pronunciation of numbers that leads to a greater digit span (Dehaene, 1997).

In our previous research we investigated whether spoken Chinese language leads to better arithmetic skills in pre-school children (Rodic et al., 2014). We assessed children from China, Russia, UK and two populations from Kyrgyzstan (Kyrgyz and Dungan), on arithmetic and other cognitive tests. As the Dungan population is ethnically similar to Chinese, speaks a form of Mandarin but uses Cyrillic (instead of character based Mandarin) as a writing system, we were able to test for the effect of the spoken language while controlling for its written aspect. Dungan children did not show any advantage in arithmetic over Kyrgyz children. This suggests that using oral Chinese, with its transparent number system and faster pronunciation of numbers, does not lead to mathematical advantage, at least for early arithmetic.

Other cognitive factors, such as spatial ability, might also play a role in the observed cross-cultural differences. For example, greater spatial complexity and increased visuo-spatial demands of Chinese reading and writing systems may lead to better mathematical performance.

Although the direction of effects and the nature of the association between spatial ability and mathematics remain unclear, they seem to be intrinsically linked. One recent genetically informative study examined the relative contribution of genetic and environmental factors to variation in spatial ability and to its relationship with different aspects of mathematics in 4174 pairs of 12-year-old twins (Tosto et al., 2014). The results suggested that, individual differences in spatial ability and different aspects of mathematics stem from both, common genetic (60%) and environmental (40%) factors. The observed correlation between spatial and mathematical ability was largely explained by overlapping genetic effects, but also overlapping environmental factors. At the level of the brain, both spatial cognition and number processing have been shown to rely on parietal lobes, especially the Intra Parietal Sulcus (Dehaene, 1997). At the behavioral level, many studies found associations between different aspects of spatial and mathematical abilities across development. For example, spatial sketchpad of working memory and mathematics performance were found to correlate (0.41) in second graders (Krajewski and Schneider, 2009). A correlation has also been observed between performance on a 3-D mental rotation task and mathematical word problem solving tasks in six graders (Van Garderen and Montague, 2003). Spatial ability has been found to correlate with mathematical ability over and above general cognitive ability in adults, both in the US (Rohde and Thompson, 2007), and China (Wei et al., 2012). Mathematically gifted adolescents perform better on spatial tasks than their non-gifted peers (Hermelin and O'Connor, 1986; Dark and Benbow, 1991).

Multiple potential mechanisms underlie the observed space-mathematics associations, from spatial representations of magnitudes on a mental number line, to spatial representations of mathematical relations, to the use of diagrams in algebraic problem solving (Geary, 1994, 1995; Hubbard et al., 2005).

It is possible that the observed advantage of representatives of East Asian cultures in mathematics can be at least partially explained by the spatial-mathematical link. Previous research indicates that spatial ability may causally contribute to mathematical learning (Rohde and Thompson, 2007; Wai et al., 2009). There is also evidence that East Asian populations show an average advantage in visuo-spatial abilities (Sakamoto and Spiers, 2014). This advantage may be related to the complexity of the character-based writing, which may either reflect or lead to superior spatial ability of some East Asian populations.

In contrast to letter-based scripts, where complexity is linear (reflected in the number of letters in a word), the complexity of Chinese characters increases with the number of elements (strokes and sub character components) packed into the same square configuration. When learning to read a Chinese character, both visual-orthographic processing and spatial analysis are essential (Tan et al., 2005). It is possible that continuous engagement in such processing leads to superior development of the relevant brain networks, which in turn leads to advantages in mathematics. In contrast, linear orthographic representations may lead to the development of the language relevant brain networks and their employment for solving mathematical problems. Support for the differential cortical number-related activity across populations was found in an fMRI study comparing native English speakers to native Chinese speakers (Tang et al., 2006). Native English speakers employed language processes for mental calculation (e.g., simple addition), while native Chinese speakers employed visuo-premotor association network for the same task.

The current study aims to investigate the spatial-mathematical link in 6–9-year-old Russian and UK children in the context of language learning over one school year. The children were assessed on a cognitive test battery measuring general skills (e.g., speed of processing), IQ, spatial ability, symbolic number understanding, non-symbolic comparison of numerosity, numerical reasoning, and arithmetic. The testing took place twice during the school year, once at the beginning and once at the end. Russian children were monolingual and began learning different second languages at the beginning of the school year. The languages included character-based systems (Chinese and Japanese) and alphabet-based systems (English and Spanish). The UK sample served as a control.

We tested the following 2 hypotheses:

- (1) Spatial ability (Mental rotation) at the beginning of the academic year will predict arithmetic scores (Simple subtraction) at the end of the academic year, in both, Russian and UK 6–9 year-old children, even after controlling for IQ scores.
- (2) Children who learn a second language that employs a spatially complex character-based writing system (Chinese and Japanese) will show a greater gain in mathematics and related abilities after 1 year of learning.

Although we are aware that the current design does not allow us to control for the effect of other linguistic factors, such as faster pronunciation of numbers and transparency of the number system, our previous study suggested no effect of the spoken language on arithmetic in the Dungan population (Rodic et al., 2014). In addition, the aim of our study is not to assess potential advantages of solving mathematical problems in a particular language (i.e., Russian children use Russian language for mathematical learning), but instead to test whether the process of learning and using spatially complex characters as a second language leads to some mathematically advantageous cognitive shift.

## Materials and Methods

### Participants

Seven hundred and twenty one 6–9 year-old children were recruited through primary schools in the UK and Russia. The children were tested in two waves, once at the beginning and once at the end of the 2012/2013 academic year. In the first wave of testing there were 155 UK participants from 5 schools in London (69 boys; mean age = 85 months, range 72–108); and 566 Russian participants from 15 schools across Russia (246 boys; mean age = 98.5, range = 88–104 months). In the second wave of testing, the number of participants has reduced to 145 UK participants (63 boys; mean age = 90 months, range = 80–105 months); and 438 Russian participants (185 boys; mean age = 105.8 months, range = 96–121 months). Attrition in the UK sample was mostly due to children changing schools. The substantial attrition in the Russian sample was largely due to a technical problem with the on-line test administration or access to remote samples in some regions.

The Russian participants were in the second year of their primary school education. Because Russian children start their primary education at 7 years of age (a year later than the UK children), they were inevitably older than the UK children of the same school year. In order to match the UK and Russian participants, both on their chronological age and years of education, half of the UK children were in the second year and half were in the third year of their primary education.

In the UK sample none of the children were learning a 2nd language at school before or during the year of testing. Although 30% of the sample (44 children) reported to be bi-lingual (indicated speaking languages other than English at home), none of the children used character-based writing systems.

All Russian children in the sample were monolingual and started learning the second language at school at the beginning of the school year. Out of 566 children, 379 started learning English language; 25 – Japanese; 74 – Spanish and English; and 88 – Chinese and English languages. On average, children had between 2 and 4 sessions (45 min each) of second language lessons per week. All schools were specialist language schools with enhanced language curricula. Selection into the language schools is not entirely random, although no special entry requirements are practiced and many children are enrolled on the basis of living proximity. However, parents' willingness to enroll children into specialist language schools and belief in the children's

ability to cope with the pressures of learning extra languages can be considered as a 'self selection' violation to random enrolment.

The project received approval from the Ethics Committees of Goldsmiths, University of London; and Tomsk State University. Parental consent was obtained prior to data collection.

### Measures and Procedure

The battery of tests included seven on-line ([www.dweipsy.com/lattice](http://www.dweipsy.com/lattice)) computerized tasks (see **Figure 1**) administered in a single session at schools. The testing lasted approximately 40 min. All tests started with practice trials and were always administered in the following order: Mental rotation, Choice reaction time, Non-symbolic comparison of numerosity, Symbolic number magnitude comparison, Simple subtraction, Number series and Raven's progressive matrices. Children indicated their responses by pressing "Q" or "P" (or corresponding Russian keys) marked with the stickers on the keyboard. For Choice reaction time, Non-symbolic comparison of numerosity and Symbolic number magnitude comparison tasks accuracy and RT (milliseconds) were recorded. For the rest of the tasks, the dependent variable was correct minus incorrect responses, correcting for guessing. The tasks are described in the following section, grouped in five categories: (1) general skills and IQ; (2) spatial ability; (3) symbolic number understanding; (4) non-symbolic number sense; (5) operating with numbers (arithmetic), and numerical reasoning. Internal validity of each measure was assessed using Cronbach's alpha analysis. The Cronbach's alphas, reported below separately for the two samples, are based on the first wave of data collection. The results from the second wave were highly similar.

### General Skills and IQ

*Choice reaction time* task (Butterworth, 2003) assessed accuracy and speed with which children responded to the dot appearing on the left (15 trials) or right (15 trials) side of the fixation '+.' The task was time-unconstrained. The inter-stimulus interval varied randomly from 1500 to 3000 ms. Cronbach's  $\alpha = 0.65$  ( $N = 154$ , UK sample) and Cronbach's  $\alpha = 0.87$  ( $N = 555$ , Russia).

*Raven's progressive matrices* (Raven et al., 1998) measured general intelligence. Participants were presented with an incomplete figure and had to identify the missing segment that would complete the figure's intrinsically regular pattern. Children used a mouse to indicate which out of the presented six segments was the correct one. The children had 4 min to go through as many trials as they could (80 trials in total). Cronbach's  $\alpha = 0.67$  ( $N = 154$ , UK sample) and Cronbach's  $\alpha = 0.73$  ( $N = 543$ , Russia).

### Spatial Ability

*Mental rotation* task (Shepard and Metzler, 1971) evaluated children's ability to mentally rotate three dimensional images. The target image was presented on the upper part of the screen, with two possible answers presented on the left and right bottom parts of the screen. The child had to decide which of the bottom two figures was matching the figure at the top by pressing either left or right button. The matching images were rotated from 15 to 345°.

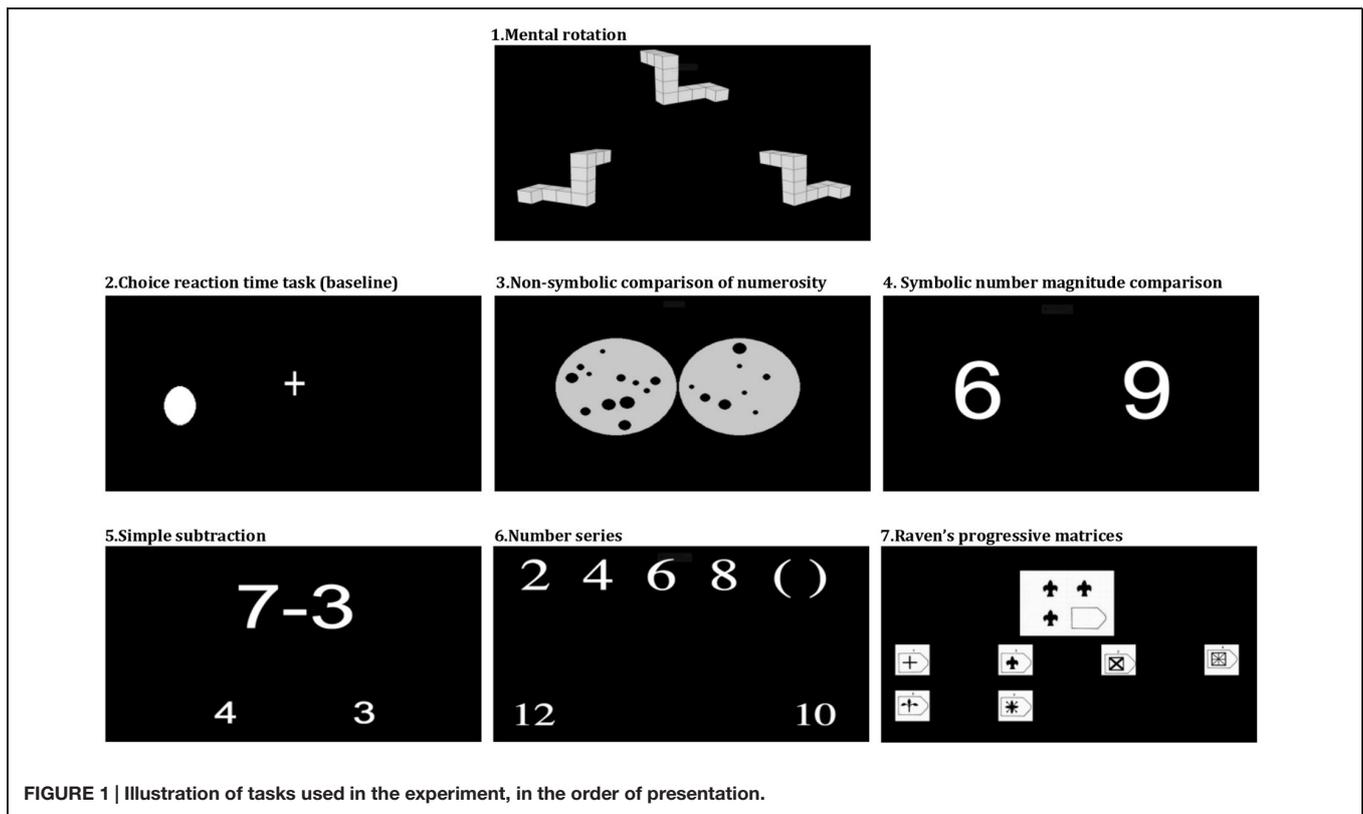


FIGURE 1 | Illustration of tasks used in the experiment, in the order of presentation.

Children had to select the correct answer in as many trials as they could in 3 min (180 trials in total). Cronbach's  $\alpha = 0.75$  ( $N = 140$ , UK sample) and Cronbach's  $\alpha = 0.87$  ( $N = 564$ , Russia).

### Symbolic Number Understanding

*Symbolic number magnitude comparison* task (Girelli et al., 2000) used a Stroop-like paradigm to assess the ability to compare numerical values of numbers. Two digits of varying sizes (1:2 size ratio) appeared simultaneously on the screen. The trials were divided into congruent, incongruent and neutral trials. In the congruent condition a numerically larger digit (e.g., 8) was also physically larger than a numerically smaller digit (e.g., 3). In the incongruent condition, three is physically larger than eight, and in the neutral condition, both digits are of the same physical size. Children had 5 s to decide which number was larger in numerical magnitude, ignoring differences in physical size. Three sessions of 28 trials each were separated by 10-s resting periods. Cronbach's  $\alpha = 0.77$  ( $N = 153$ , UK sample) and Cronbach's  $\alpha = 0.87$  ( $N = 545$ , Russia).

### Non-Symbolic Number Sense

*Non-symbolic comparison of numerosity* (Baroody and Ginsburg, 1990) measured non-symbolic number sense. Children had to estimate (without counting) which of the two sets of dots of varying sizes, presented simultaneously on the screen, contained more dots (36 trials, 5 s per trial). In all sets the combined area of all dots was controlled to be the same. The number of dots varied from 5 to 12; ratios were 2:3, 5:7, and 3:4. Cronbach's  $\alpha = 0.78$

( $N = 153$ , UK sample) and Cronbach's  $\alpha = 0.84$  ( $N = 549$ , Russia).

### Operating with Numbers (arithmetic) and Numerical Reasoning

*Simple subtraction* task assessed early arithmetic ability. The minuends were all smaller than 18 and the differences were single-digit numbers. Two candidate answers were presented beneath the problem, one on each side of the screen. Children had to select the correct answer in as many trials as they could in 2 min (92 problems). Correct and incorrect answers were within the range of each other plus or minus 3. Cronbach's  $\alpha = 0.75$  ( $N = 152$ , UK sample) and Cronbach's  $\alpha = 0.73$  ( $N = 542$ , Russia).

*Number series completion* task (Smith et al., 2001) measured logical numerical reasoning. A sequence of numbers was presented on the screen (e.g., 1,3,5,7) with two additional numbers below it. The child was asked to infer the pattern of these numbers and decide which out of the two candidate answers presented below the sequence should complete the sequence (e.g., 9 or 16). The children were given 4 min to do as many sequences as they could. Cronbach's  $\alpha = 0.65$  ( $N = 146$ , UK sample) and Cronbach's  $\alpha = 0.63$  ( $N = 589$ , Russia).

## Results

### Growth

First, we evaluated average growth on each assessed measure over one academic year. This was done separately for the UK and

Russian samples, as the two samples could not be directly compared: UK sample was heterogeneous in terms of biological age and years of schooling; the Russian children were selected from specialist language schools (see **Table 1** for mean and SDs for the raw scores for both samples on all tasks).

As can be seen from **Table 2**, children's performance improved significantly for all tasks, with the exception of RT in Choice Reaction Time in the UK sample. The effect sizes of growth, obtained by means of one way repeated measures ANOVAs, ranged from 2.1% (for RT and accuracy of Choice RT task in the Russian sample) to 44% (for Simple subtraction in the UK sample).

Further, we ran the between-subjects one-way ANOVAs on growth scores for each variable, calculated by subtracting the scores at time 1 from the scores at time 2, with sample as a two level factor (UK vs. Russian). The size of growth for all variables was highly similar across the Russian and the UK samples, with only one significant [but negligible,  $\eta_p^2 = 1.2\%$ ;  $F(1,525) = 6.161$ ,  $p = 0.01$ ] difference for the Raven's task (see **Table 2**).

## The Relationship between Spatial Ability and Arithmetic Over Time

The cross-lag analyses, conducted on each sample separately, tested the first hypothesis regarding the longitudinal relationship between spatial ability and arithmetic, while controlling for IQ scores. This type of analysis (described below) evaluates associations between the two variables over time, while controlling for stability of each measure over time and for associations between the two measures at the same time.

### Russian Sample

Before conducting the cross-lagged analyses, a correlation matrix was obtained and inspected to check for longitudinal associations, as well as associations between Mental rotation and Subtraction. Correlations between time 1 and time 2 assessments were moderate, both for Mental rotation ( $r = 0.507$ ) and Subtraction ( $r = 0.498$ ), indicating relative stability of measures over time. A modest relationship between the Mental rotation and Subtraction was found at both assessments waves ( $r = 0.221$  at time 1 and  $r = 0.275$  at time 2). Correlation between Mental rotation at time 1 and Subtraction at time 2 was slightly higher ( $r = 0.277$ ) than that of Subtraction at time 1 and Mental rotation at time 2 ( $r = 0.137$ ).

Next, the cross-lag structural equation modeling (Campbell, 1963), was utilized to investigate the longitudinal relationship between spatial ability (Mental rotation) and early arithmetic (Subtraction). This type of analysis can investigate causal ordering of variables by estimating three types of relationships: (1) autoregressive paths which assess within-construct stability by estimating the correlation between two assessments of the same variable (e.g., Mental rotation at time 1 and time 2); (2) contemporaneous relationship between the two measures at the same assessment wave (e.g., Mental rotation at time 1 and Subtraction at time 1); and (3) cross-lagged relationship which estimates the extent to which scores for one variable at time 1 predict unique variance in the other variable at a later time (e.g., Mental rotation at time 1 and Subtraction at time 2), while controlling for

autoregressive and contemporaneous associations. Further, we included the Raven's scores at time 1 as a covariate in order to control for IQ on both measures at both times.

**Figure 2** shows standardized path coefficients for the longitudinal relationship between spatial ability and arithmetic. The full model, which included the cross-lagged associations, was found to fit the data better (AIC = 4825.61), than the model excluding those associations (AIC = 4837.52). The non-significant paths were then dropped from the cross-lagged model until the best fitting model was achieved:  $\chi^2(3) = 6.35$ ,  $p = 0.098$ , RMSEA = 0.046, CFI = 0.990, TLI = 0.966, SRMR = 0.024 ( $N = 527$ ). The best fitting model suggests the direction of the relationship from spatial ability to later arithmetic and not vice versa. The standardized paths are shown in **Figure 2**. Significant paths were: the cross-lagged path from Mental rotation at time 1 to Subtraction at time 2 ( $\beta = 0.180$ , SE = 0.04,  $p < 0.001$ ); the contemporaneous paths between Mental rotation and Subtraction at both, time 1 ( $\beta = 0.225$ , SE = 0.04,  $p < 0.001$ ) and time 2 ( $\beta = 0.162$ , SE = 0.04,  $p = 0.002$ ); and the autoregressive paths for both, Mental rotation ( $\beta = 0.524$ , SE = 0.04,  $p < 0.001$ ) and arithmetic ( $\beta = 0.526$ , SE = 0.04,  $p < 0.001$ ). The paths from the covariate (Raven's) were significant for Mental rotation at time 1 ( $\beta = 0.103$ , SE = 0.04,  $p = 0.018$ ); and Subtraction at time 1 ( $\beta = 0.136$ , SE = 0.04,  $p = 0.002$ ).

### The UK Sample

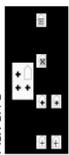
Correlations between time 1 and time 2 assessments were moderate, both for Mental rotation ( $r = 0.434$ ) and Subtraction ( $r = 0.575$ ), indicating relative stability of measures over time. A modest relationship between the Mental rotation and Subtraction was found at both assessments waves ( $r = 0.185$  at time 1 and  $r = 0.195$  at time 2). Correlation between Mental rotation at time 1 and Subtraction at time 2 was not significant, while Subtraction at time 1 and Mental rotation at time 2 were modestly correlated ( $r = 0.209$ ).

Next, cross-lag analysis was conducted in order to investigate the relationship between the spatial ability (Mental rotation) and early arithmetic (Subtraction) while accounting for the IQ scores.

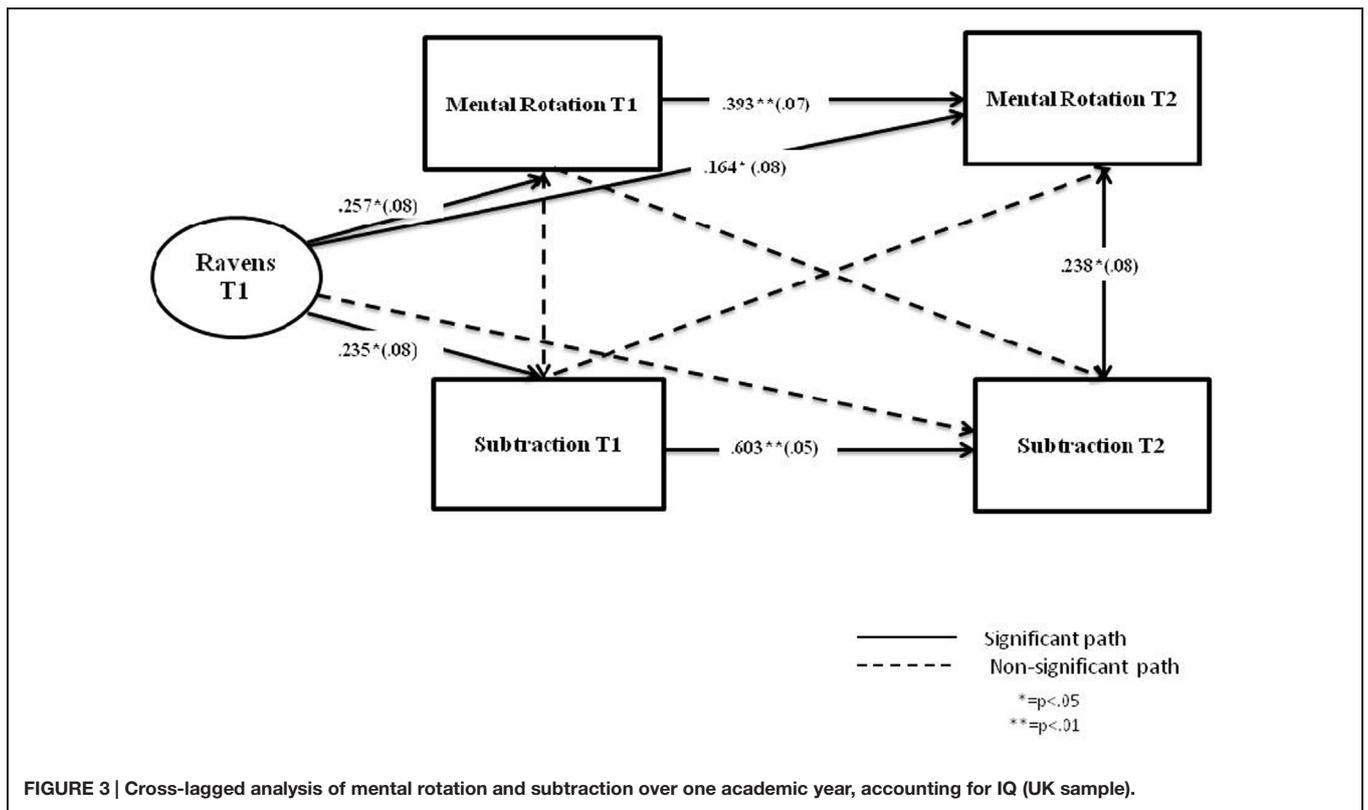
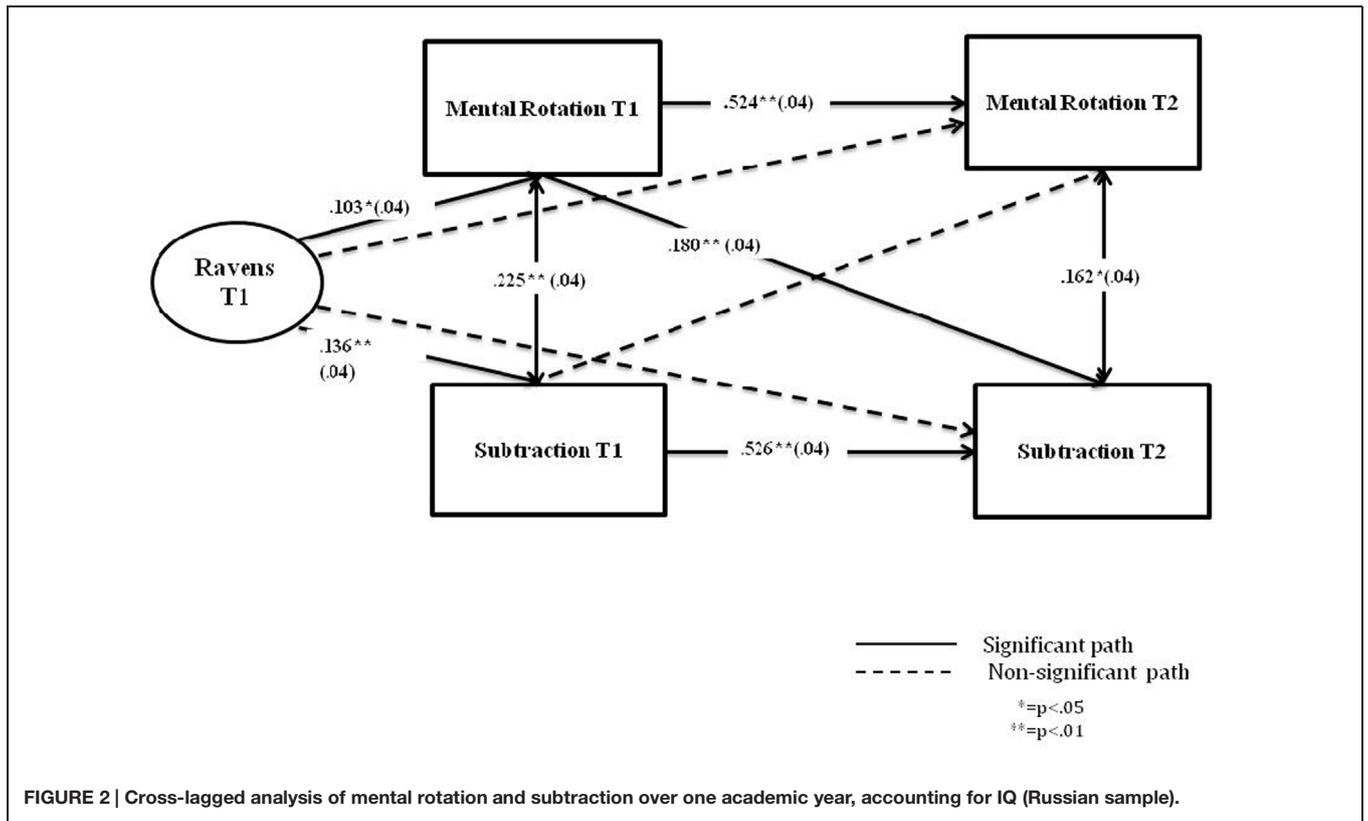
**Figure 3** shows standardized path coefficients for the longitudinal relationship between spatial ability and arithmetic in the UK sample. The model excluding the cross-lag associations was found to fit the data better (AIC = 1428.342) than the full model which included those associations (AIC = 1425.781). The non-significant paths were then dropped from the model. The model in **Figure 3** fitted the data very well:  $\chi^2(4) = 2.618$ ,  $p = 0.062$ ; RMSEA < 0.001; CFI = 1.00; TLI = 1.031; SRMR = 0.036 ( $N = 144$ ). In the UK sample, the significant paths included: the contemporaneous path between Mental rotation and Subtraction at time 2 ( $\beta = 0.238$ , SE = 0.08,  $p = 0.004$ ); and the autoregressive paths for both Mental rotation ( $\beta = 0.393$ , SE = 0.07,  $p < 0.001$ ) and Subtraction ( $\beta = 0.603$ , SE = 0.05,  $p < 0.001$ ).

The paths from the covariate (Raven's) were significant for Mental rotation at time 1 ( $\beta = 0.257$ , SE = 0.08,  $p = 0.001$ ); Mental rotation at time 2 ( $\beta = 0.164$ , SE = 0.08,  $p = 0.035$ ); and Subtraction at time 1 ( $\beta = 0.235$ , SE = 0.08,  $p = 0.003$ ).

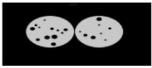
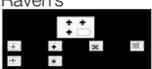
TABLE 1 | Descriptive statistics for UK and Russian samples, for all tasks at Time 1 and Time 2.

Task	Sample	UK T1	Russia T1	UK T2	Russia T2
	Score type (max. score overall)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
Mental rotation 	Correct-incorrect responses (43)	8.9 (8.12) N = 155	12.74 (9.28) N = 544	11.69 (9.76) N = 145	14.80 (11.22) N = 438
Choice RT 	Mean RT (2500)	651.20 (168.49) N = 152	651.60 (232.12) N = 556	620.49 (145.23) N = 143	595.61 (156.36) N = 437
	Proportion of correct responses (1)	0.96 (0.06) N = 152	0.94 (0.12) N = 556	0.97 (0.03) N = 145	0.95 (0.08) N = 556
Comparison of numerosity 	Mean RT (4940)	1356.84 (358.69) N = 151	1187.13 (337.30) N = 551	1228.70 (280.16) N = 151	1110.32 (431.72) N = 431
	Proportion of correct responses (1)	0.86 (0.11) N = 151	0.86 (0.13) N = 551	0.91 (0.08) N = 145	0.89 (0.11) N = 364
Magnitude comparison 	Mean RT (2434)	1306.5 (290.46) N = 150	1118.80 (266.09) N = 547	1161.69 (245.35) N = 144	985.16 (210.69) N = 438
	Proportion of correct responses (1)	0.89 (0.06) N = 150	0.88 (0.09) N = 551	0.92 (0.05) N = 145	0.91 (0.08) N = 438
Subtraction 	Correct-incorrect responses (50)	9.8 (10.00) N = 155	15.87 (10.33) N = 579	16.66 (8.79) N = 144	22.75 (10.03) N = 431
Number series 	Correct-incorrect responses (20)	4.11 (5.30) N = 149	2.8 (6.65) N = 537	7.02 (4.90) N = 144	4.81 (6.77) N = 431
Raven's 	Correct-incorrect responses (33)	7.3 (6.5) N = 148	11.84 (5.36) N = 517	11.24 (6.70) N = 142	14.29 (4.46) N = 431

UK T1, average performance of the UK sample in the first wave of testing (Time 1); UK T2, average performance of the UK sample in the second wave of testing (Time 2); Russia T1, average performance of the Russian sample in the first wave of testing (Time 1); Russia T2, average performance of the Russian sample in the second wave of testing (Time 2).



**TABLE 2 | The effect sizes for growth over the school year, on all tasks for the UK and Russian samples.**

	Sample	UK	Russia
Task	Score type	Partial $\eta^2$ (%)	Partial $\eta^2$ (%)
Mental rotation 	Correct–incorrect responses	7%** $N = 155$	6.1%** $N = 544$
Choice RT 	Mean RT	non-significant $N = 152$	2.1%* $N = 556$
	Proportion of correct responses	3.4%* $N = 152$	2.1%* $N = 556$
Comparison of numerosity 	Mean RT	13%** $N = 151$	4.2%** $N = 551$
	Proportion of correct responses (1)	11%** $N = 151$	3.1%** $N = 551$
Magnitude comparison 	Mean RT	26%** $N = 150$	35%** $N = 547$
	Proportion of correct responses	16%** $N = 150$	6.3%** $N = 551$
Subtraction 	Correct–incorrect responses	44%** $N = 155$	30%** $N = 579$
Number series 	Correct–incorrect responses	15%** $N = 149$	4.2%** $N = 537$
Raven's 	Correct–incorrect responses	22%** $N = 148$	11%** $N = 517$

\* $p < 0.05$ , \*\* $p < 0.001$  for significance of improvement in scores (growth) from Time 1 to Time 2; ns, non-significant; NS, number series.

## Second Language Acquisition Effects on Cognitive Skills and Arithmetic

The second hypothesis, addressing the effects of second language learning on arithmetic and related skills, was investigated in the Russian sample. The sample was split into four groups based on the different languages that children learn at school (i.e., English, Japanese, Spanish and English and Chinese and English). **Table 3** shows the descriptive statistics for the four language groups at both times.

One-way ANOVAs were employed to test for differences on all tasks between the four groups at the beginning of the year (time 1). Despite the differences in sample sizes (379 for English; 25 for Japanese; 74 for Spanish; 88 for Chinese), performance at baseline was overall similar across the four groups. No significant differences between the groups were found for RT and accuracy of Choice reaction time; RT and accuracy of Symbolic number magnitude comparison; RT and accuracy of Non-symbolic comparison of numerosity; and Mental rotation (correct minus incorrect responses score). For the remaining three tasks significant ( $p < 0.05$ ), but very small ( $\eta_p^2 = 2.1\text{--}3.2\%$ ,  $p < 0.05$ ) differences were found (details available from the authors). Only

one violation to equal variance was found (for the Raven's task), but the differences in variance were negligible (as suggested in Field, 2009).

Next, to test for the effect of the language learnt at school on task performance at the end of the year, we conducted ANCOVA, including the performance on the task at the baseline time 1 as a covariate. The only significant effect of language was on the Number series completion task [ $F(3,399) = 4.063$ ,  $p = 0.007$ ], with Chinese/English learning group slightly ( $\eta_p^2 = 3\%$ ), outperforming Japanese ( $p = 0.021$ ) and Spanish/English ( $p = 0.001$ ) learning groups.

## Discussion

The study set out to investigate the relationship between spatial ability and mathematical performance. First, we assessed the cross-lag relationship between spatial ability (Mental rotation) and arithmetic (Simple subtraction). The significant positive link from spatial ability to later arithmetic was found only in the Russian sample. This finding is similar to what was previously

TABLE 3 | Descriptive statistics for the Russian sample language groups, for all tasks at both times.

Task	Score type (maximum score overall)	English T1	Spanish and English T1	Chinese and English T1	Japanese T1	English T2	Spanish and English T2	Chinese and English T2	Japanese T2
		Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
Mental rotation 	Correct-incorrect responses (43)	12.9 (9.35) N = 366	11.06 (9.63) N = 70	13.77 (8.67) N = 83	11.28 (9.13) N = 25	14.27 (11.63) N = 290	14.53 (11.38) N = 54	17.17 (9.38) N = 76	14.11 (10.85) N = 18
	Mean RT (2500)	646.81 (253.36) N = 369	627.46 (0.11) N = 71	680.99 (231.17) N = 88	698.36 (196.21) N = 25	596.67 (165.38) N = 289	570.84 (143.00) N = 55	587.85 (129.09) N = 76	647.63 (147.32) N = 17
Choice RT 	Proportion of correct responses (1)	0.93 (0.12) N = 369	0.94 (0.11) N = 71	0.93 (0.14) N = 88	0.96 (0.06) N = 25	0.95 (0.08) N = 289	0.95 (0.09) N = 55	0.96 (0.06) N = 76	0.96 (0.04) N = 18
	Mean RT (2562)	1186.09 (327.03) N = 365	1190.82 (351.27) N = 71	1160.50 (290.90) N = 88	1308.44 (535.40) N = 25	1155.02 (495.88) N = 217	975.73 (21.76) N = 52	1029.30 (258.04) N = 78	1130.06 (178.23) N = 18
Comparison of numerosity 	Proportion of correct responses (1)	0.87 (0.13) N = 365	0.86 (0.13) N = 71	0.86 (0.10) N = 88	0.82 (0.21) N = 25	0.89 (0.12) N = 217	0.89 (0.09) N = 52	0.88 (0.11) N = 77	0.92 (0.04) N = 18
	Mean RT (2218)	1110.39 (272.96) N = 363	1151.93 (273.29) N = 69	1110.84 (238.54) N = 88	1207.04 (204.27) N = 25	981.35 (219.64) N = 286	964.92 (172.16) N = 55	990.15 (202.38) N = 79	1085.53 (194.62) N = 18
Magnitude comparison 	Proportion of correct responses(1)	0.88 (0.08) N = 363	0.86 (0.09) N = 69	0.87 (0.09) N = 88	0.89 (0.07) N = 25	0.91 (0.08) N = 286	0.89 (0.08) N = 55	0.92 (0.06) N = 79	0.91 (0.09) N = 18
	Correct-incorrect responses (50)	16.65 (10.88) N = 391	13.38 (8.72) N = 74	16.12 (9.13) N = 88	10.88 (7.07) N = 25	23.16 (10.73) N = 284	20.75 (7.6) N = 55	23.43 (9.3) N = 74	19.66 (6.76) N = 18
Subtraction 	Correct-incorrect responses (17)	3.37 (6.38) N = 358	0.77 (6.79) N = 70	2.59 (7.20) N = 83	0.84 (6.87) N = 25	4.96 (6.82) N = 282	2.24 (6.80) N = 54	6.47 (6.36) N = 78	2.94 (5.40) N = 17
	Mean RT (2218)	1110.39 (272.96) N = 363	1151.93 (273.29) N = 69	1110.84 (238.54) N = 88	1207.04 (204.27) N = 25	981.35 (219.64) N = 286	964.92 (172.16) N = 55	990.15 (202.38) N = 79	1085.53 (194.62) N = 18
NS 	Correct-incorrect responses (29)	11.97 (5.23) N = 349	10.40 (3.66) N = 71	12.93 (6.94) N = 88	10.33 (2.69) N = 24	4.72 (6.82) N = 282	13.92 (4.15) N = 52	14.93 (2.79) N = 78	12.94 (3.38) N = 18
	Mean RT (2218)	1110.39 (272.96) N = 363	1151.93 (273.29) N = 69	1110.84 (238.54) N = 88	1207.04 (204.27) N = 25	981.35 (219.64) N = 286	964.92 (172.16) N = 55	990.15 (202.38) N = 79	1085.53 (194.62) N = 18

English T1; Spanish and English T1; and Japanese T1, average performance of children before learning a second language in the first wave of testing (Time 1); English T2; Spanish and English T2; Chinese and English T2; and Japanese T2, average performance of children after learning a second language for one academic year in the second wave of testing (Time 2).

found with 18-year-old students from the US whose spatial ability predicted mathematical portion of Scholastic Aptitude Tests (SATs) even after controlling for IQ (Rohde and Thompson, 2007). The smaller number of participants and the larger standard errors in the UK sample indicated possibility of insufficient power to detect any effects that might have existed in this sample. Further, as spatial ability is a complex multifactorial domain, our findings may not extend beyond the relationship between 3-D mental rotation ability and arithmetic. Future studies with tasks measuring different aspects of spatial ability (e.g., spatial memory or navigation) are needed to assess whether different aspects of spatial ability have different relation to arithmetic.

Differences between the two samples may also indicate that the relationship between spatial ability and mathematical ability may develop differently in different cultures. However, as the UK children were between 12 and 15 months younger than the Russian children at both waves of testing, the differences could also reflect developmental processes. Future research is needed to confirm the generalizability of our finding to different populations and at different ages.

Second, owing to the access of special sampling of the Russian sample we were able to investigate whether acquisition and usage of the character-based writing system, within 1 year, could lead to a better performance in arithmetic and other cognitive skills in 6–9-year children. There were no noticeable differences between the language groups on any of the tasks at the baseline time 1. At time 2 no significant differences in performance emerged across the language groups for most tasks. The only task that showed significant, although small effect (3% of the variance) was the Number series completion task, even after controlling for the performance at the baseline. The children who learnt Chinese/English showed a small advantage over those who learnt Japanese and Spanish/English. Because these children showed the biggest improvement in this mathematically related task over 1 year, there is some indication that learning the Chinese language may positively influence mathematical reasoning. As the children were learning mathematics in Russian and were not tested in Chinese, oral advantage of Chinese language is unlikely to explain the observed advantage. It is possible that the usage of the spatially complex character-based writing system indeed plays a role in the observed advantage in mathematical ability, as suggested by our hypothesis. The lack of advantage in children who learnt Japanese, which also required learning the character-based writing system, could be due to the very small sample size of this sample ( $N = 25$ ), but further research is needed in order to test this.

Overall, there was a significant improvement over one academic year, in both samples on all tasks. The biggest improvement in both samples (Russian = 30% and the UK = 44%), was seen for the Simple subtraction task. This is not surprising as this ability was explicitly taught to the pupils throughout the year. Furthermore, although the UK sample demonstrated bigger growth on most tasks overall, the only significant difference between samples was found on Raven's task, with UK children showing bigger growth. This finding suggests that the developmental trajectory of mathematically relevant skills is similar for both samples.

The writing system is likely to be only one of many factors contributing to the advantage of East Asian children in mathematics. As discussed earlier, cultural ethos, parental support, frequent practice and the Confucian values that place high value on effort and academic success (Leung, 2001) – may all contribute independently.

Another possible explanation for the lack of the effect of learning a spatially complex character-based writing system is that our sample was too young. Previous studies suggested that mathematical advantage in children with better spatial skills was due to them employing spatial representations to solve mathematical problems (Geary, 1994, 1995). This skill comes with more experience and might not be used by the children in early primary school. In order for children to employ such strategies, more mathematical experience and explicit teaching of these strategies may be needed. Investigations with older children are required to explore these possibilities. Additionally, the usage of spatial strategies might not be useful for simple arithmetic. For the advantage in arithmetic in this early stage other factors might play a bigger role, such as the fast pronunciation of numbers (Geary et al., 1993) and regularity of number systems. Using spatial strategies as suggested by Geary (1995) might begin to play a role with more advanced mathematical problem solving and geometry, which is taught later in formal education.

The study had several limitations. Several tasks lacked sensitivity as they turned out to be too easy (e.g., Non-symbolic comparison of numerosity) or too difficult (e.g., Number series completion). It is possible that more sensitive tasks would yield some significant differences between the language groups. Further research with more sensitive tasks is needed to better address these issues.

Another limitation, is that the length of the period in which children were exposed to learning the character-based writing system might not have been sufficient enough for the development of greater spatial and, consequently, greater mathematical skills. It is likely that a longer and more intensive exposure (more than 3 h a week) is needed for an effect to emerge. In addition, also it is of course possible that the mathematical advantage of Asian populations is not influenced by the usage of character-based writing system, but reflects a particular distinctive cognitive feature that has led to the invention of this complex system in the first place.

Finally, the sample sizes of our language groups were very different, ranging from 25 participants for the Japanese language group to 391 participants for the English language group. The fact that Japanese language group consisted of only 25 participants at time 1 and 18 at time 2 could have significantly decreased a chance of detecting any true effects of learning that language.

## Conclusion

Despite curricular and other sample differences, the rate of learning on all tasks over one academic year was very similar for

the UK and Russian children. In line with previous research, spatial ability predicted arithmetic in the Russian sample longitudinally and beyond intelligence scores. We extended previous literature by testing whether the acquisition of spatially complex character-based writing system could lead to better performance in maths, due to the established relationship between the spatial ability and mathematics. Only a small effect (3%) of learning Chinese as a second language was found on mathematical reasoning. Our findings suggest

that despite the importance of spatial ability for mathematics, one academic year of increased spatial processing through exposure to spatially complex writing systems might not be enough to provide a mathematical advantage. Longer periods of exposure might be needed for it to have a positive effect on mathematics. Further cross-cultural longitudinal research is needed to identify specific cognitive, cultural, educational, linguistic and genetic influences on mathematical learning.

## References

- Baroody, A. J., and Ginsburg, H. P. (1990). "Children's mathematical learning: a cognitive view," in *Constructivist Views on Teaching and Learning Mathematics, Journal for Research in Mathematical Education Monograph*, eds R. B. Davis, C. A. Maher, and N. Noddings (Reston, VA: National Council of Teaching of Mathematics), 79–90.
- Butterworth, B. (2003). *Dyscalculia Screener*. Gosport: Ashford Colour Press.
- Campbell, D. T. (1963). "From description to experimentation: interpreting trends in quasi experiments," in *Problems in Measuring Change*, ed. C. W. Harris (Madison, WI: University of Wisconsin Press).
- Dark, V. J., and Benbow, C. P. (1991). Differential enhancement of working memory with mathematical and vertical precocity. *J. Educ. Psychol.* 83, 48–60. doi: 10.1037/0022-0663.83.1.48
- Dehaene, S. (1997). *The Number Sense: How the Mind Creates Mathematics*. New York, NY: Oxford University Press.
- Field, A. P. (2009). *Discovering Statistics Using SPSS: and Sex and Drugs and Rock 'n' Roll*, 3 Edn. London: Sage publications.
- Geary, D. C. (1994). *Children's Mathematical Development: Research and Practical Applications*. Washington, DC: American Psychological Association.
- Geary, D. C. (1995). Reflection of evolution and culture in children's cognition. *Am. Psychol.* 50, 24–37. doi: 10.1037/0003-066X.50.1.24
- Geary, D. C., Bow-Thomas, C. C., Fan, L., and Siegler, R. S. (1993). Even before formal instruction, Chinese children outperform American children in mental addition. *Cogn. Dev.* 8, 517–529. doi: 10.1016/S0885-2014(05)80007-3
- Girelli, L., Lucangel, D., and Butterworth, B. (2000). The development of automaticity in accessing number magnitude. *J. Exp. Child Psychol.* 76, 104–122. doi: 10.1006/jecp.2000.2564
- Hermelin, B., and O'Connor, N. (1986). Spatial representations in mathematically and in artistically gifted children. *Br. J. Educ. Psychol.* 56, 150–157. doi: 10.1111/j.2044-8279.1986.tb02656.x
- Hubbard, T., Andrews, D., Caccamo, M., Cameron, G., Chen, Y., Clamp, M., et al. (2005). Ensembl 2005. *Nucleic acid res.* 33, D447–D453. doi: 10.1093/nar/gki138
- Imbo, I., and Vandierendonck, A. (2007). The development of strategy use in elementary school children: working memory and individual differences. *J. Exp. Child Psychol.* 96, 284–309. doi: 10.1016/j.jecp.2006.09.001
- Krajewski, K., and Schneider, W. (2009). Exploring the impact of phonological awareness, visuospatial working memory, and preschool quantity-number competencies on mathematics achievement in elementary school: findings from a 3-year longitudinal study. *J. Exp. Child Psychol.* 103, 516–531. doi: 10.1016/j.jecp.2009.03.009
- Leung, C. (2001). English as an additional language: distinctive language focus or diffused curriculum concerns? *Lang. Educ.* 15, 33–55.
- Miura, I. T. (1987). Mathematics achievement as a function of language. *J. Educ. Psychol.* 79, 79–82. doi: 10.1037/0022-0663.79.1.79
- Mullis, I. V. S., Martin, M. O., and Foy, P. (2008). *TIMSS 2007 International Mathematics Report: Findings from IEA's Trends in International Mathematics and Science Study at Eight and Fourth Grades*. Chestnut Hill, MA: TIMSS and PIRLS International Study Center, Boston College.
- OECD. (2010). *PISA 2009 Results: Executive Summary*. Washington, DC: U.S. Government Printing Office.
- Raven, J., Raven, J. C., and Court, J. H. (1998). *Manual for Raven's Progressive Matrices and Vocabulary Scales*. Oxford: Oxford Psychologists Press
- Rodić, M., Zhou, X., Tikhomirova, T., Wei, W., Malykh, S., Ismatulina, V., et al. (2014). Cross-cultural investigation into cognitive underpinnings of individual differences in early arithmetic. *Dev. Sci.* 18, 1–10.
- Rohde, T. E., and Thompson, L. A. (2007). Predicting academic achievement with cognitive ability. *Intelligence* 35, 83–92. doi: 10.1016/j.intell.2006.05.004
- Sakamoto, M., and Spiers, M. V. (2014). Sex and cultural differences in spatial performance between Japanese and North Americans. *Arch. Sex. Behav.* 43, 483–491. doi: 10.1007/s10508-013-0232-8
- Shepard, R., and Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science* 171, 701–703. doi: 10.1126/science.171.3972.701
- Smith, P., Fernandes, C., and Strand, S. (2001). *Cognitive Abilities Test: Technical Manual*, 3rd Edn. London: nferNelson.
- Song, M. J., and Ginsburg, H. P. (1988). The effect of the Korean number system on young children's counting: a natural experiment in numerical bilingualism. *Int. J. Psychol.* 23, 319–332. doi: 10.1080/00207598808247769
- Stevenson, H. W., Chen, C., and Lee, S. Y. (1993). Mathematics achievement of Chinese, Japanese, and American children: ten years later. *Science* 259, 53–58. doi: 10.1126/science.8418494
- Stevenson, H. W., and Stigler, J. W. (1992). *The Learning Gap*. New York, NY: Free Press.
- Tan, L. H., Spinks, J. A., Eden, G., Perfetti, C. A., and Siok, W. T. (2005). Reading depends on writing, in Chinese. *Proc. Natl. Acad. Sci. U.S.A.* 102, 8781–8785. doi: 10.1073/pnas.0503523102
- Tang, Y., Zhang, W., Chen, K., Feng, S., Ji, Y., Shen, J., et al. (2006). Arithmetic processing in brain shaped by cultures. *Proc. Natl. Acad. Sci. U.S.A.* 103, 10775–10780. doi: 10.1073/pnas.0604416103
- Tosto, M., Hanscombe, K. B., Haworth, C. M. A., Davis, O. S. P., Petrill, S. A., Dale, P. S., et al. (2014). Why do spatial abilities predict mathematical performance? *Dev. Sci.* 17, 1–9. doi: 10.1111/desc.12138
- Van Garderen, D., and Montague, M. (2003). Visual-spatial representation, mathematical problem solving, and students of varying abilities. *Learn. Disabil. Res. Pract.* 18, 246–254. doi: 10.1111/1540-5826.00079
- Wai, J., Lubinski, D., and Benbow, C. P. (2009). Spatial ability for STEM domains: aligning over 50 years of cumulative psychological knowledge solidifies its importance. *J. Educ. Psychol.* 101, 817–835. doi: 10.1037/a0016127
- Wei, W., Yuan, H., Chen, C., and Zhou, X. (2012). Cognitive correlates of performance in advanced mathematics. *Br. J. Educ. Psychol.* 82, 157–181. doi: 10.1111/j.2044-8279.2011.02049.x

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Rodić, Tikhomirova, Koliienko, Malykh, Bogdanova, Zueva, Gynku, Wan, Zhou and Kovas. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Mathematics and reading difficulty subtypes: minor phonological influences on mathematics for 5–7-years-old

Julie A. Jordan\*, Judith Wylie and Gerry Mulhern

School of Education, Queen's University Belfast, Belfast, UK

## Edited by:

Ann Dowker, University of Oxford, UK

## Reviewed by:

Michele M. Mazzocco, University of Minnesota, USA

Hannah Pimperton, University College London, UK

Steve Chinn, Independent Researcher, UK

Tuire Katriina Koponen, University of Jyväskylä, Finland

## \*Correspondence:

Julie A. Jordan, School of Education, Queen's University Belfast, 69-71 University Street, Belfast BT7 1HL, UK  
e-mail: ja.jordan@qub.ac.uk

Linguistic influences in mathematics have previously been explored through subtyping methodology and by taking advantage of the componential nature of mathematics and variations in language requirements that exist across tasks. The present longitudinal investigation aimed to examine the language requirements of mathematical tasks in young children aged 5–7 years. Initially, 256 children were screened for mathematics and reading difficulties (RDs) using standardized measures. Those scoring at or below the 35th percentile on either dimension were classified as having difficulty. From this screening, 115 children were allocated to each of the mathematical difficulty (MD;  $n = 26$ ), MDRD ( $n = 32$ ), RD ( $n = 22$ ) and typically achieving ( $n = 35$ ) subtypes. These children were tested at four time points, separated by 6 monthly intervals, on a battery of seven mathematical tasks. Growth curve analysis indicated that, in contrast to previous research on older children, young children with MD and MDRD had very similar patterns of development on all mathematical tasks. Overall, the subtype comparisons suggested that language played only a minor mediating role in most tasks, and this was secondary in importance to non-verbal skills. Correlational evidence suggested that children from the different subtypes could have been using different mixes of verbal and non-verbal strategies to solve the mathematical problems.

**Keywords:** subtyping, language, mathematical difficulties, children, longitudinal, reading

## INTRODUCTION

A variety of methodologies have shed light on the nature of the relationship between language and mathematics including cross-cultural, correlational, and neuroscientific approaches (e.g., Butterworth, 2008; Dowker et al., 2008). One approach is to compare the mathematics performance of children with different levels of academic achievement, with a focus on subtype differences that mimic the subgroups of children who are grouped in classrooms on the basis of their ability level (e.g., Geary and Hoard, 2001; Koponen et al., 2006; Donlan et al., 2007). In a longitudinal study of children aged 7–9 years adopting both a componential and subtyping approach, Hanich et al. (2001) and Jordan et al. (2003) reported that children with specific mathematical difficulties (MDs) had an advantage over those with comorbid mathematics and reading difficulties (MDRD) in areas where performance may be mediated by language, specifically exact calculation, story problems, and calculation principles. On the other hand, these groups did not differ on tasks reliant on numerical magnitudes, visuo-spatial processing, or automaticity, such as approximate arithmetic. Of course, the curriculum changes as children progress through school and becomes progressively more language dominated, meaning that the relationship between language and mathematics cannot be assumed to be static.

Using a subtyping approach, the present research examined the language requirements of Hanich et al. (2001) and Jordan et al. (2003) mathematical tasks for younger children aged 5–7 years. In contrast to N. Jordan and colleagues' research on older children, standardized reading tests would not have been

suitable for the younger children in the present research. Therefore classifications in the present research were made based on phonological ability, which is strongly associated with early reading progress (Adams, 1990; Ziegler et al., 2010) and with specific language difficulty (e.g., Kamhi and Catts, 1986; Catts et al., 2005). For simplicity, in this paper, the term RD is used to represent both reading difficulty (RD) and phonological difficulty. Inferences about the role of language in mathematics were made by comparing the performance of four subtypes: specific MDs; specific phonological difficulties (RD), comorbid mathematics and phonological difficulties (MDRD) and typical mathematics and phonological achievement (TA). Consistent with Hanich et al. (2001) and Jordan et al. (2003) these subtypes were compared on seven mathematical tasks; namely, exact calculation; story problems, approximate arithmetic, place value, calculation principles, forced retrieval, and written problems.

Hanich et al. (2001) and Jordan et al. (2003) made their conclusions about the language requirements of the tasks based on comparisons between MD and MDRD. They concluded that there was little evidence of MDs amongst RD relative to TA. In contrast, the value of RD/TA comparisons has been demonstrated by Jordan et al. (2010) who found that amongst RD children who did not have MD at age 5 years, approximately half had standardized mathematical ability consistent with MDRD by age 7 years. Closer examination revealed that this was due to the age-related shift in balance from non-verbal to verbal mathematical items in the standardized mathematics achievement test. Indeed, RD made less progress than TA on the more verbal tasks such as number facts,

formal calculation, and formal concepts, but had similar growth on tasks with lower language requirements including numbering, number comparison, and informal concepts. As both MD/MDRD and RD/TA subtype comparisons can tell us about the importance of language in mathematical tasks, the present research focuses on both. Further, building upon the work of previous subtyping studies (e.g., Hanich et al., 2001; Jordan et al., 2003), the present research evaluated subtyping as an approach to examining the role of language in mathematics. For this reason the possibility that the relationship between language and mathematical tasks is obscured by subtypes adopting different compensatory strategies is explored. Hereafter follows a synopsis of what is currently known about the language requirements of these seven mathematical tasks.

Exact calculation is an untimed task involving questions such as “how much is 3 plus 5?” or “how much is 6 take away 3?” Previous studies have suggested that language skills are unique predictors of performance on this task (Swanson and Beebe-Frankenberger, 2004; Fuchs et al., 2005, 2006). A longitudinal study examining the mathematical abilities of 5–9-year-old children with specific language impairment (SLI) suggests that these counting-related skills are indeed verbally mediated. The key problem areas identified at age five in these children included producing the number word sequence and counting accurately (Fazio, 1999). Hanich et al. (2001) found that 7-year-old children with MDRD had a more severe impairment in exact calculation than those with MD only. The advantage of MD over MDRD on this task appears to be due to MD’s more accurate use of verbal/finger counting procedures and comparatively better understanding of calculation principles (Jordan and Montani, 1997; Geary et al., 1999; Jordan and Hanich, 2000). Clearly there is strong evidence to suggest this task is verbally demanding for young children, and these effects can be observed from as young as 5 years. Although children with MD were found to outperform MDRD on this task, they still did not perform as well as typically achieving (TA) children at age 7 (Hanich et al., 2001), which is unsurprising given the verbal and non-verbal requirements of counting (Dowker, 2005).

Story problems are untimed arithmetic problems presented in word format that rely on both verbal and non-verbal abilities (Swanson and Beebe-Frankenberger, 2004; Fuchs et al., 2006), and the language requirements of this task are considerably greater than those of exact calculation. Good language skills will help the children to understand the meaning of the story problem, to subsequently form a problem representation, and to read and review the problem rather than relying on holding the problem in memory. Indeed, Jordan et al. (1995) had previously found that children aged 6 with low language ability but adequate spatial ability were impaired on this task relative to normally achieving children. Of course, other non-linguistic skills are also important such as the ability to form concrete or numerical representations of word problems (Dowker, 2005). Subtyping evidence highlights the importance of language ability for this task; comparisons of mathematical subtypes showed that children aged 7–9 years with MDRD consistently perform less well on story problems than those with MD (Hanich et al., 2001; Jordan et al., 2003). Hanich et al. (2001) suggested that, although the performance of MD was weakened by their mathematical deficits, such children may have been

able to compensate, to an extent, through their unimpaired verbal skills, and therefore outperform MDRD. Likewise the unimpaired mathematical skills of the RD subtype may have helped alleviate the negative impact of their poor language skills when performing this task. By contrast, the difficulties observed in MDRD, who have weaknesses in both mathematics and reading may have been due their limited compensatory skills. These ideas are speculative and the exact nature of compensatory routes to problem solving is unclear. It is perhaps surprising that the RD subtype did not display a stronger impairment on this task, because understanding the problem through language has been highlighted as a particular area of difficulty for children.

A distinction between approximate (e.g.,  $2 + 3 = 4$  or 11) and exact (e.g.,  $2 + 3 = ?$ ) arithmetic has been made in educational research (Dowker, 2003). Despite sharing some key skills (e.g., using relations between numbers) and performance on these tasks being associated in young children (Dowker, 1998), discrepancies and dissociations have been found between these tasks in typically developing children (Dowker, 1994, 1998), neuropsychological patients with dyscalculia (Warrington, 1982; Dehaene and Cohen, 1991), and adults with dyslexia (Gobel and Snowling, 2010). Cross-cultural research highlights that cultures that lack number words beyond 5 are able to perform approximate but not exact arithmetic when the problems involve numbers outside their vocabulary range (Pica et al., 2004). Imaging studies show that exact calculation produces greater activation of areas of the brain associated with language, while performing approximate arithmetic leads to greater activation of areas involved in the processing of quantity and spatial information (Dehaene et al., 1999). Subtyping evidence based on 7–9-year-old also indicates that approximate arithmetic has relatively low language demands; both MD and MDRD displayed a similar level of impairment, while RD performed as well as TA (Hanich et al., 2001; Jordan et al., 2003).

Place value tasks assess understanding of how the position of a digit represents a value, as well as ability to name numbers. Children who speak a language with a regular counting system such as Welsh are better at reading two digit numbers than those who speak English which has an irregular counting system (Dowker et al., 2008). Correlational evidence shows that linguistic skills are related to performance on a number naming task, as is spatial span but to a lesser extent than linguistic ability (LeFevre et al., 2010). Subtyping studies indicate that children with MD outperform MDRD on this task (Jordan and Hanich, 2000), and those with RD (Hanich et al., 2001) and SLI (Grauberg, 1998) have difficulty compared to normally achieving children. Contrary to this idea, Hanich et al. (2001) reported that MD and MDRD had a similar level of performance on a place value task. They also found that both MD and MDRD were impaired relative to TA children, concluding that non-verbal skills must also be important. Jordan et al. (2003) found little difference between the subtypes on number naming, suggesting that this part of the task was too easy for children aged 7–9 years, although it is likely that differences will be found in younger children. Overall these findings indicate that both verbal and non-verbal abilities facilitate performance on this task.

Calculation principles such as commutativity,  $n + 1$  and inversion can be used by children to infer the answers to mathematics problems rather than having to fully calculate the answer. Dowker (1998) found that for children aged 5–9 years verbal IQ predicts the use of calculation principles on addition tasks, while both verbal and performance IQ are predictive for subtraction; also predictive of calculation principles use on addition tasks was a verbal/performance IQ discrepancy, possibly because uneven abilities make it difficult to follow standard school-taught procedures, leading children to adopt alternative strategies. Hanich et al. (2001) and Jordan et al. (2003) proposed that when these principles are taught at school, language comprehension may be key to developing a conceptual understanding of them. Subtyping studies have shown that at age 7 children with MD performed at the same level as MDRD; however, by age 9 children with MD significantly outperformed MDRD (Hanich et al., 2001; Jordan et al., 2003).

Fact retrieval assesses the ability to recall answers to problems directly from memory. Subtyping evidence indicates that poor fact retrieval is the most consistent deficit in children with MDs (Russell and Ginsburg, 1984; Geary, 1990, 1993; Geary et al., 1991; Barrouillet et al., 1997; Ostad, 1997, 1998, 1999, 2000; Hanich et al., 2001; Jordan et al., 2003) and in individuals with Turner syndrome who have normal reading ability (Rovet et al., 1994; Molko et al., 2003; Bruandet et al., 2004). These findings strongly indicate that non-verbal factors must influence performance on this task. Although fact retrieval deficits have been identified as a defining feature of MD by many studies, care must be taken when interpreting this finding. As Dowker (2004) points out, arithmetic screening tests often emphasize fact retrieval, consequently it is unsurprising that those children identified as MD on the basis of that test display impairments on a fact retrieval task. While non-verbal skills such as subitizing ability appear to facilitate performance on forced retrieval tasks (Koontz and Berch, 1996), language is also important, as children and adults with specific RDs do not perform as well as normally achieving children on forced retrieval (Geary et al., 2000; Hanich et al., 2001; Simmons and Singleton, 2006; Smedt and Boets, 2010), nor do children with SLIs (Fazio, 1999). There are a number of reasons why children with RDs experience fact retrieval difficulties. For example, Robinson et al. (2002) point out that the repetition method of learning mathematical facts relies very heavily on phonological ability. Additionally, counting is a verbally mediated skill which is commonly used by young children to solve arithmetic problems and correctly solving these problems through counting will strengthen the association between the problem and the solution (Siegler and Shrager, 1984).

Written problems are presented in a vertical visual format and are not read to the children (e.g., Hanich et al., 2001; Jordan et al., 2003). As all problems are displayed in vertical format it is inevitable that some degree of spatial ability is needed for the correct placement and alignment of digits (Dowker, 2005). Evidence suggesting that this task requires good non-verbal skills comes from a study of children with visuo-spatial learning difficulty but normal reading ability (Venneri et al., 2003). Despite performing similar to controls on an oral calculation task, these children displayed impairments on a written calculation task. In addition, Hanich et al. (2001) and Jordan et al. (2003) found that both subtypes with MD had a similar level of impairment on this

task, and those with specific RDs did not. This indicates that non-verbal ability plays a greater role than verbal ability in this task. The written problems task used by Jordan et al. (2003) involved problems both with and without a carry/borrow operation. As items with carry/borrow operations are not included in the curriculum for the age group involved in the present study, these items are not included in our adapted version of this task. Relative to normally achieving children, those with visuo-spatial learning difficulty have more difficulty when a carry/borrow operation is required than when it is not (Venneri et al., 2003). Therefore, by removing this requirement, the task makes fewer non-verbal demands and this must be taken into consideration when making predictions about the performance of the subtypes on this task.

Our predictions about the role of language in each of the seven mathematical tasks were made based on studies of older children with MD and what we already know about the normal development of children aged 5–7 years. It is expected that subtyping evidence will indicate that both verbal and non-verbal skills are important for tasks such as exact calculation, story problems, calculation principles, place value, and forced retrieval. On the other hand, performance on tasks such as written problems and approximate arithmetic is likely to involve relatively fewer language skills. In some ways language could play a more important role in task performance in the early years because children aged 5–7 years are more reliant on verbal counting-based procedures than older children (Siegler, 1996). It is possible, however, that as the language skills of the children in the present research will be less well-developed than the sample in Hanich et al. (2001) and Jordan et al. (2003), the TA children will not yet have developed as much of an advantage. Since the maths curriculum becomes progressively more language dominated over the early school years, the relation between language and mathematics cannot be assumed to be static. In this study we explore the consistency of MD and RD relationships in the earliest school years, in children 5–7 years of age.

## MATERIALS AND METHODS

### PARTICIPANTS

The 14 participating schools in this study were from a range of demographic areas, including representation from both urban and rural areas. The Northern Ireland Multiple Deprivation Measure (Northern Ireland Statistics and Research Agency, 2005) rankings for each school's intake area (1 highest, 890 lowest), indicated that about half of the schools in the sample were located in deprived areas and the other half in the more affluent areas of Northern Ireland (range 2–887). All Year 1 children in the participating schools who had parental consent took part in the screening exercise. The mathematics and phonological difficulty screening tests were individually administered to 256 children with a typical testing session lasting 25–30 min. All participants spoke English as their first language. From this screening, 115 children were retained to allow for comparable sample sizes in the four subtypes of interest (see **Table 1**). At the time of screening the children were aged 5½ years ( $M = 65.59$  months;  $SD = 3.61$ ), and slightly more males (55%) took part than females.

The specific achievement criteria for each subtype are as follows:

**Table 1 | Subtype ability characteristics and sample sizes.**

Subtype	N				Mathematical percentile score Mean (SD)	Phonology percentile score Mean (SD)	Verbal percentile score Mean (SD)	Non-verbal percentile score Mean (SD)
	Time							
	1	2	3	4				
MDRD	32	29	30	29	21.34 (9.44)	20.98 (11.57)	22.80 (16.35)	37.72 (23.04)
MD	26	25	25	24	24.42 (10.89)	46.96 (19.21)	42.62 (19.32)	38.81 (27.05)
RD	22	24	19	20	49.27 (14.37)	21.82 (10.76)	31.90 (20.32)	48.62 (25.93)
TA	35	33	29	29	53.57 (16.29)	54.93 (13.63)	46.72 (20.93)	44.63 (23.45)

MD: Mathematics score at or below the 35th percentile, and phonological score at or above the 40th percentile.

RD: Phonological score at or below the 35th percentile, and mathematics score at or above the 40th percentile.

MDRD: Both mathematics and phonological scores at or below the 35th percentile.

TA: Both mathematics and phonological scores at or above the 40th percentile.

None: Children with phonological/mathematics scores within the 36th–39th percentile range were unclassified.

### SCREENING MEASURES

Standardized mathematics ability: the Test of Early Mathematics Ability 3, Form A (TEMA 3, Ginsburg and Baroody, 2003) was designed to identify young children with MDs aged 3:0–8:11 years. This test examines formal and informal mathematical skills including number comparison, non-verbal arithmetic, counting, problem solving, numbering skills, numeral literacy, mastery of number facts, calculation skills, and the understanding of concepts. In a study by Mazzocco and Myers (2003) which employed various standardized tests, the Test of Early Mathematics Ability, TEMA-2 (Ginsburg and Baroody, 1990) was reported as the test which produced the most normally distributed data and the greatest stability in test performance over time. The TEMA-3 test has high test–retest reliability (0.95) and correlates moderately (0.55) with the applied problems subtest of the Woodcock–Johnson III Tests of Achievement (Woodcock et al., 2001).

Standardized phonological ability: the Rhyme Detection and Phoneme Deletion (beginning sounds) subtests of the Phonological Abilities Test (PAT; Muter et al., 1997) measure young children's phonological ability, which is a strong predictor of early reading progress (Adams, 1990). The Rhyme Detection subtest requires a child to select which of three words rhyme with the stimulus word (e.g., cat, which word rhymes?, fish, gun, or hat). For the Phoneme Deletion (beginning sounds) subtest the child is required to delete the first phoneme of a single syllable word (e.g., “bus” without the [b] says [us]). The Rhyme Detection and Phoneme Deletion – Beginning Sounds subtests were selected because overall they are considered to be the best predictors at age 5, 6, and 7 years of scores on the BAS word reading test (Elliott et al., 1997), and they have good test–retest reliability (Phoneme Deletion, 0.84; Rhyme Detection, 0.80).

### VERBAL AND NON-VERBAL ABILITY MEASURES

The Verbal cluster (Word Definitions and Verbal Similarities) and the Non-Verbal subscale (Matrices) of the British Ability Scales 2 (BAS-2; Elliott et al., 1997) were used as ability measures at time 2. In the word definitions test children were presented orally with a word and asked what it meant. In order to be scored as correct, the child had to express the key concepts of the word's meaning, rather than simply to use it in the correct context. The Verbal Similarities test assesses a child's ability to explain how two words are similar. For example, when asked why an apple and orange are alike they could say they are both fruits. More general answers that would apply to other categories (e.g., both have skins) are scored as incorrect. The purpose of the matrices subtest is to examine a child's ability to correctly identify those rules that govern variables in abstract figures. For each item the child must choose which of six alternatives correspond to the geometric pattern that is missing from the matrix. The verbal cluster has a correlation of 0.69 with the corresponding scale of the WISC III, and the non-verbal reasoning cluster has a correlation of 0.56 with the performance scale of the WISC III. All subtests have good internal reliability for 6-years-old (word definitions, 0.79; verbal similarities, 0.88; matrices, 0.78).

### BATTERY OF MATHEMATICAL TASKS

The mathematics test battery comprised seven tasks: exact calculation, story problems, approximate arithmetic, place value, calculation principles, forced retrieval, and written problems. These tasks were closely based on those used previously by N. Jordan and colleagues with 7–9-years-old. A number of adjustments were made to the tasks so that they would be suitable for children aged 5–7 years. (1) The time limits for approximate arithmetic, calculation principles, and forced retrieval tasks were increased to accommodate the slower processing speeds typical of younger children. (2) The administration time of N. Jordan's battery was considered too long for young children and therefore the number of items in each task was reduced for the present investigation. (3) Digit correspondence items were omitted from the place value task as they were considered to be too difficult for children aged 5–7 years. (4) Problems with a carry operation were excluded from the written problems task, because this concept is not taught during the early years of primary school. These tasks are described in further detail in Jordan et al. (2009).

## PROCEDURE

**Table 1** displays the ability information for each subtype in the experimental sample, and sample sizes at each time the mathematical test battery was administered. From the 256 children screened, 115 were allocated to the four achievement subtypes and completed the mathematical tasks at time 1. Attrition rates for times 2, 3, and 4 were 3, 10, and 11% respectively. This total sample of 115 included all children identified as having MD or RD. There were too many MDRD and TA children to retain for further longitudinal testing from the 256 children screened. Therefore a subset of children with MDRD was kept; these children were selected carefully to ensure that MDRD were well-matched to MD for mathematics ability and to RD for phonological ability. Similarly, TA children were selected to match the MD group for phonological ability and the RD group for mathematics ability.

All testing was completed on an individual basis at the participating schools by one experimenter who had received police clearance. The study was approved by the School of Psychology Research Ethics Committee at Queen's University Belfast. The children from the four achievement subtypes were assessed longitudinally on a battery of mathematical tasks from age 5½ years onwards. Each child completed the mathematical test battery at four time points separated by 6 monthly intervals, and the administration duration for each session was on average 25 min. Four versions of the battery were constructed in which the order of items was varied for the exact calculation, story problems, approximate arithmetic, and forced retrieval tasks. Each child was given a different version of the test battery at the four time points; the presentation order across the four time points for these versions was varied within each subtype. For all children, the tasks were presented in the following order, (1) exact calculation, (2) story problems, (3) approximate arithmetic, (4) place value, (5) calculation principles, (6) forced retrieval, and (7) written problems. The verbal and non-verbal ability measures were administered at age 6–106 of the 115 (9 were absent) participating children. Testing took 20–30 min depending on the ability level of the child.

## RESULTS

### DATA ANALYSIS PROCEDURES

Raw mean scores and standard deviations are shown in **Table 2**, while estimated trajectories are shown in **Figure 1**. All models were estimated by maximum likelihood (ML) using AMOS 7 (Arbuckle, 2006). Prior to the data analysis, individual and group level growth plots for each of the mathematical subtasks were examined; these provided an indication of the approximate shape of growth for each task. These plots revealed that, for all subtypes, growth appeared to be approximately linear on story problems, approximate arithmetic, place value, forced retrieval and written problems tasks, and curvilinear on exact calculation and calculation principles tasks. It was also apparent that for all tasks there was considerable variation in final status and to a lesser extent growth rates, not only between, but also within, subtypes.

Data analysis consisted of two stages, the first of which involved fitting an unconditional model (without predictors) for the whole sample to each of the seven mathematical tasks, to determine if linear or non-linear models provided better fit. In the second

stage of the analysis, conditional models were fit to each mathematical task, with achievement group membership as a predictor. Three types of model were tested in this analysis including, linear, freed loading, and quadratic. For all models the slope loading for the fourth time point was set to 0, in order to scale the intercept factor to represent final status. For both linear and non-linear models, the measurement occasions were parameterised in such a way as to reflect rates of growth in terms of 6-months increments.

### LINEAR AND NON-LINEAR UNCONDITIONAL MODEL COMPARISONS

For all tasks, nested model comparisons were used to evaluate whether growth was linear or non-linear. Chi-square difference tests were used to evaluate if the specification of a freed loading model provided a significantly better model fit than a linear model. The results indicated that a non-linear model did not significantly improve model fit for five of the tasks (story problems, approximate arithmetic, place value, forced retrieval, and written problems) suggesting that growth for these tasks was probably linear. By contrast, the chi-square difference test was significant for the exact calculation ( $\chi^2 = 13.47$ ,  $df = 2$ ,  $p < 0.01$ ) and for the calculation principles task ( $\chi^2 = 13.04$ ,  $df = 2$ ,  $p < 0.01$ ). This would suggest that a non-linear model would better describe the shape of growth for these tasks.

When a quadratic model was run for the calculation principles task multiple estimation problems were encountered, which, according to Bollen and Curran (2006) suggests that this model provides a poor representation of the observed data. In such cases where growth does not follow a strict linear or quadratic trajectory a freed loading model is more suitable, therefore a freed loading model was specified for the calculation principles task. On the other hand, the quadratic model did provide a good fit for the exact calculation task. Although the mean of this factor ( $\chi^2 = 9.673$ ,  $df = 1$ ,  $p < 0.01$ ) was significantly different from 0, the variance was not. As there was little variation in acceleration then there would be no value in using achievement subtype membership as a predictor. It would still have been possible to use a quadratic model for this task by fixing the variance; however, to provide more comparability in terms of the interpretation of growth rates across tasks, a freed loading model was also specified for this task.

According to the chi-square test statistics all the models fit well, as there was no significant difference between the models and the data (**Table 3**). The model for story problems and calculation principles do not provide an exact fit according to the root-mean-square error of approximation (RMSEA) statistics; nevertheless, these values are still considered acceptable (Browne and Cudeck, 1993). All models fit well-according to the Tucker Lewis index (TLI) and incremental fit index (IFI) statistics (between 0.9 and 1.2).

**Table 4** displays the means and variances for final status and the growth rates for the combined sample on each task. For all tasks the variances for the growth rates and final status were significantly greater than zero, therefore the analysis of parameter correlates could be pursued. In the next stage of data analysis, achievement subtype was added as a predictor to the model for each task.

**Table 2 | Mean raw scores and standard deviation on the mathematical tasks by subtype at times 1–4.**

Task	Subtype	Time 1		Time 2		Time 3		Time 4	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD
Exact calculation	MDRD	0.81	1.15	1.66	1.63	2.90	1.79	3.76	1.86
	MD	1.62	1.30	3.24	2.03	4.04	1.77	4.63	1.74
	RD	1.82	1.47	3.79	1.89	4.79	1.44	5.20	1.24
	TA	2.74	1.72	4.12	1.73	4.97	1.32	5.59	0.68
Story problems	MDRD	0.84	0.85	1.41	1.09	2.00	1.36	2.72	1.60
	MD	1.46	1.36	2.16	1.57	2.36	1.78	3.33	1.90
	RD	1.14	0.94	2.63	1.50	3.32	1.63	4.25	1.41
	TA	2.17	1.25	2.82	1.74	4.10	1.52	4.66	1.74
Approximate arithmetic	MDRD	5.91	2.43	7.41	2.10	8.23	2.22	8.83	2.07
	MD	6.85	2.39	8.16	2.17	8.44	2.77	9.96	1.78
	RD	7.09	2.64	8.04	2.85	9.16	1.57	10.40	1.85
	TA	7.46	2.23	8.30	2.53	9.31	2.22	10.55	2.03
Place value	MDRD	1.84	0.81	2.45	0.74	2.80	0.76	3.41	0.98
	MD	2.69	0.79	2.84	1.14	3.64	0.91	4.17	0.76
	RD	2.23	0.61	3.00	0.59	3.47	0.70	4.15	0.88
	TA	2.83	1.07	3.55	0.97	4.34	1.26	5.00	1.13
Calculation principles	MDRD	0.13	0.42	0.10	0.41	0.93	1.01	1.55	1.43
	MD	0.42	0.76	0.88	1.09	2.28	1.59	2.54	1.67
	RD	0.32	0.65	1.08	1.06	2.26	1.66	3.55	1.70
	TA	1.11	1.30	1.76	1.44	3.28	1.79	4.00	1.71
Forced retrieval	MDRD	0.63	0.87	0.55	0.83	1.70	1.73	2.59	1.97
	MD	1.19	1.13	2.16	1.65	3.40	1.76	3.58	1.47
	RD	1.23	1.02	2.13	1.54	3.37	1.71	4.10	2.10
	TA	2.00	1.33	3.24	1.73	4.07	1.60	5.07	0.84
Written problems	MDRD	0.31	0.54	1.34	1.54	2.13	2.03	3.31	2.35
	MD	1.08	1.87	2.28	2.03	3.20	2.63	5.25	2.95
	RD	1.18	1.01	3.00	2.36	4.32	2.58	5.35	2.62
	TA	1.77	1.66	3.76	2.45	4.72	2.67	6.00	1.91

Maximum possible score by task: exact calculation (6), story problems (8) approximate arithmetic (13), place value (7), calculation principles (6); forced retrieval (6); written problems (8).

**CONDITIONAL MODELS WITH ACHIEVEMENT GROUP MEMBERSHIP AS A PREDICTOR**

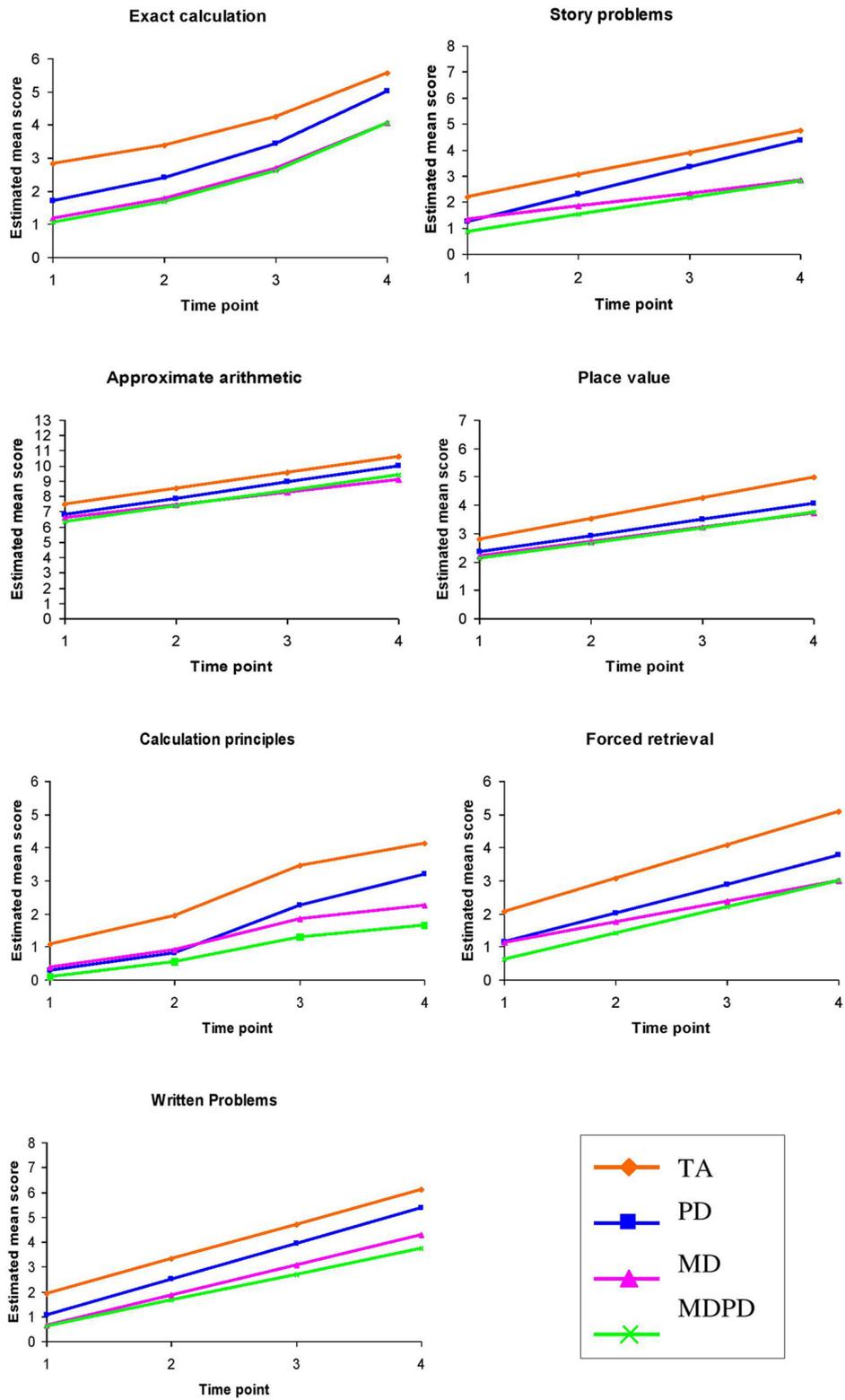
To enable between-group comparisons, final status and growth rates were regressed on three dummy variables. In the first set of models, MD, RD, and TA were coded as 1 and MDRD, the reference group, was coded as 0. In order to compare all groups, models were also estimated with TA and then with RD as the reference group.

The fit indices (Table 5), show that most models still fit well after the predictor was added and the model fit actually improved for the story problems and calculation principles tasks. The fit indices for the approximate arithmetic task model are not as good as they were before achievement subtype was added to the model; despite this the overall model fit for this task is still acceptable.

For all tasks, there was significant variation in final status which was unexplained by achievement subtype membership

(Table 6). With the exception of story problems, after controlling for achievement subtype membership, there was still considerable unexplained variance in growth rates. In fact, for all tasks, achievement subtype membership explained much less of the variance in growth rate than in final status. Achievement subtype membership explained much more variance in the growth rates for story problems (24%) and calculation principles (19%) than for the other mathematical tasks. From the remaining tasks, approximate arithmetic is the one for which achievement subtype membership explains the least variance, both in terms of final status (12%) and growth rates (2%). It is likely that, for these reasons, the model for the approximate arithmetic task fits less well after achievement subtype membership was added as a predictor to the model.

Growth curve model comparisons between the MD and MDRD subtypes revealed no significant differences in terms of final status



Note. Time 1 = 5 ½ years; Time 2 = 6 years; Time 3 = 6 ½ years; Time 4 = 7 years

FIGURE 1 | Estimated mean scores on the mathematical tasks by achievement subtype.

**Table 3 | Fit indices for the final unconditional models.**

Task	$\chi^2$	TLI	IFI	RMSEA
Exact calculation	$p = 0.42$	1.01	1.00	0.00
Story problems	$p = 0.18$	0.97	0.98	0.07
Approximate arithmetic	$p = 0.90$	1.16	1.07	0.00
Place value	$p = 0.63$	1.03	1.01	0.00
Calculation principles	$p = 0.15$	0.94	0.98	0.08
Forced retrieval	$p = 0.58$	1.02	1.01	0.00
Written problems	$p = 0.66$	1.03	1.01	0.00

Fit indices Ideal fit; Chi-square test statistic ( $\chi^2$ ) = non-significant p-value; Tucker-Lewis index (TLI) = 1; Incremental fit index (IFI) = 1; Root-mean-square error of approximation (RMSEA)  $\leq 0.05$ .

**Table 4 | Estimated parameters for the combined sample by task.**

Task	Final status		Growth rate		Covariance (FS/GR)
	Mean	Variance	Mean	Variance	
Exact calculation	4.71	2.19	0.98	0.25	0.36
Story problems	3.72	2.87	0.75	0.16	0.58
Approximate arithmetic	9.83	2.92	0.99	0.36	0.70
Place value	4.17	1.01	0.59	0.09	0.23
Calculation principles	2.82	2.83	0.77	0.24	0.73
Forced retrieval	3.76	3.18	0.83	0.29	0.76
Written problems	4.89	5.33	1.26	0.43	1.21

FS/GR is final status/growth rate. All significant at the  $p < 0.05$  level.

and growth rates on any of the mathematical tasks (Table 7 and Figure 1). Furthermore, both subtypes had significantly lower final status on all tasks relative to TA children. The MD subtype displayed significantly weaker growth over the 18 months period than normally achieving children on the story problems, place value, calculation principles and forced retrieval tasks. Despite

**Table 5 | Fit indices for the conditional models.**

Task	$\chi^2$	TLI	IFI	RMSEA
Exact calculation	$p = 0.454$	1.00	1.00	0.00
Story Problems	$p = 0.193$	0.96	0.99	0.05
Approximate arithmetic	$p = 0.072$	0.85	0.95	0.08
Place value	$p = 0.322$	0.98	0.99	0.04
Calculation principles	$p = 0.423$	1.00	1.00	0.01
Forced retrieval	$p = 0.258$	0.97	1.00	0.05
Written problems	$p = 0.287$	0.97	0.99	0.04

**Table 6 | Variance explained by achievement subtype membership.**

Task	Final status		Growth rate	
	Variance	R <sup>2</sup>	Variance	R <sup>2</sup>
Exact calculation	1.72**	0.20	0.23**	0.02
Story problems	2.03**	0.29	0.12	0.24
Approximate arithmetic	2.59**	0.12	0.38*	0.02
Place value	0.73**	0.29	0.09**	0.08
Calculation principles	1.84**	0.36	0.21**	0.19
Forced retrieval	2.37**	0.26	0.26**	0.07
Written problems	4.28**	0.18	0.36**	0.07

Variance refers to the variance in intercepts and slopes remaining after controlling for achievement subtype membership. R<sup>2</sup> the amount of variance in the model explained by achievement subtype membership. \*  $p < 0.05$ , \*\*  $p < 0.01$ .

MDRD and MD having similar growth rates across tasks, the only task on which MDRD experienced significantly less growth than normally achieving children was calculation principles.

The RD subtype had significantly greater final status than both MD and MDRD on the exact calculation and story problems tasks and only the MDRD subtype on calculation principles. On the story problems and calculation principles tasks the RD subtype had significantly greater growth than both the MD and MDRD subtypes.

**Table 7 | Estimated final status (age 7 years) and growth rates by achievement subtype.**

	MDRD		MD		RD		TA	
	FS	GR	FS	GR	FS	GR	FS	GR
Exact calculation	4.07 <sup>a,b</sup>	1.00	4.06 <sup>a,b</sup>	0.96	5.02	1.10	5.56	0.91
Story problems	2.80 <sup>a,b</sup>	0.64 <sup>b</sup>	2.82 <sup>a,b</sup>	0.50 <sup>a,b</sup>	4.38	1.05	4.74	0.84
Approximate arithmetic	9.39 <sup>a</sup>	1.01	9.12 <sup>a</sup>	0.84	10.00	1.06	10.63	1.05
Place value	3.74 <sup>a</sup>	0.54	3.72 <sup>a</sup>	0.50 <sup>a</sup>	4.06 <sup>a</sup>	0.57	4.98	0.73
Calculation principles	1.66 <sup>a,b</sup>	0.52 <sup>a,b</sup>	2.26 <sup>a</sup>	0.63 <sup>a,b</sup>	3.19 <sup>a</sup>	0.97	4.13	1.02
Forced retrieval	2.99 <sup>a</sup>	0.79	2.99 <sup>a</sup>	0.62 <sup>a</sup>	3.76 <sup>a</sup>	0.87	5.09	1.01
Written problems	3.73 <sup>a</sup>	1.04	4.30 <sup>a</sup>	1.22	5.39	1.44	6.10	1.39

FS (final status), GR (growth rate). Significant differences,  $p < 0.05$ . <sup>a</sup>TA > MDRD, MD, RD; <sup>b</sup>RD > MDRD, MD.

**Table 8 | Correlations between phonological, verbal and non-verbal ability, and performance on each mathematical task by subtype.**

Subtype	Task	Phonological ability	Verbal ability	Non-verbal ability
TA	Exact calculation	0.11	-0.31*	-0.03
	Story problems	0.12	0.43*	0.23
	Approximate arithmetic	0.34*	0.17	-0.03
	Place value	0.41*	0.40*	0.40*
	Calculation principles	0.04	0.15	0.29
	Forced retrieval	0.05	-0.09	0.11
	Written problems	0.40*	-0.01	0.05
RD	Exact calculation	-0.25	-0.15	0.33
	Story problems	0.03	0.10	0.46*
	Approximate arithmetic	-0.12	0.19	0.60*
	Place value	-0.26	0.15	0.42*
	Calculation principles	0.01	0.32	0.46*
	Forced retrieval	-0.01	-0.21	0.43*
	Written problems	-0.14	-0.03	0.46*
MD	Exact calculation	0.78*	0.03	0.38*
	Story problems	0.76*	0.39*	0.38*
	Approximate arithmetic	0.63*	-0.03	0.13
	Place value	0.59*	0.21	0.22
	Calculation principles	0.46*	0.17	0.33
	Forced retrieval	0.87*	0.13	0.28
	Written problems	0.72*	0.22	0.23
MDRD	Exact calculation	0.21	-0.06	0.37*
	Story problems	0.40*	-0.06	0.34*
	Approximate arithmetic	0.19	0.15	-0.05
	Place value	0.18	0.02	0.36*
	Calculation principles	0.16	0.28	0.07
	Forced retrieval	0.28	-0.07	0.58*
	Written problems	0.10	-0.06	0.34*

\* $p < 0.05$ .

Children with specific RDs performed less well than normally achieving children at time 4 on all tasks; these differences were significant for place value, calculation principles, and forced retrieval. Despite these differences, RD and TA had comparable growth rates across all tasks. Ceiling effects were apparent on exact calculation and forced retrieval for the normally achieving subtype at the end of the developmental period under investigation. Consequently, these effects may have impeded our ability to detect significant differences between the subtypes with learning difficulties and the TA subtype in terms of final status and growth rate on these tasks. Based on the estimated scores

produced by the growth curve analysis, overall the consistent pattern for all tasks (Figure 1) was: TA outperformed RD, and MD and MDRD had a similar level of impairment relative to RD and TA.

#### RELATIONSHIPS BETWEEN VERBAL, NON-VERBAL ABILITY, AND THE MATHEMATICAL TASKS

The relationship between verbal, non-verbal and phonological ability and performance on each of the mathematical tasks (time 4) was investigated using Pearson product-moment correlations. Scores on the ability measures were correlated with performance on each mathematical task to examine the relationship between these abilities in TA children and in the subtypes with learning difficulties (Table 8).

#### DISCUSSION

The present research examined the role of phonological ability in the mathematical development of 5–7-years-old using a subtyping approach. Contrary to Hanich et al. (2001) and Jordan et al. (2003), both MD and MDRD children aged 5–7 years in the present study exhibited very similar performance across all mathematical tasks, as evidenced by their final status (age 7 years) and growth rates. Despite initial matching for mathematics ability with TA, RD had consistently weaker performance on place value, calculation principles, and forced retrieval, suggesting that phonological ability is important for children aged 5–7 years when performing these particular tasks. In addition to age-related differences, some of the adaptations made to N. Jordan's original battery of tasks may have led to minor qualitative differences in the nature of the tasks, possibly limiting comparability with the present investigation. Furthermore, the use of different mathematics and RD screening may partly explain the differences in findings between the present research and that of Hanich et al. (2001) and Jordan et al. (2003). While phonological ability is related to both language and reading ability, as Robinson et al. (2002) point out, phonological ability may directly influence mathematics achievement. For example, the repetition method of learning mathematical facts relies very heavily on phonological ability. As each number fact is repeated phonological information must be both generated and stored and each repetition strengthens the association between the problem and the answer. The greater the association between the answer and the problem the greater the chance of successful recall. This may explain why children with poor phonological ability but strong non-verbal abilities were more impaired in the present research compared to children with specific RD in other research (Hanich et al., 2001; Jordan et al., 2003).

As MD and MDRD were initially matched for mathematics ability, it was not expected that MDRD would perform worse than MD on all tasks. Rather it was expected that MDRD would have weaker performance than MD on tasks with stronger language requirements, and have similar or possibly better results than MD on tasks with fewer language requirements if they could adopt effective compensatory strategies. Despite a body of research showing that language plays a key role in many of the mathematical tasks, the MD and MDRD subtypes performed similarly on all tasks. It is difficult to explain why RD performed worse than TA on some tasks, yet MDRD and MD had

similar performance despite having different phonological abilities. Of course not all skills associated with mathematics were assessed in this study and it is possible that MDRD were able to achieve comparable performance to MD through the use of alternative skills. Indeed, uncertainty exists over the exact number of deficits that may contribute to children's MDs (Swanson, 2007) and to what extent these occur in isolation or co-occur in various combinations. To date, numerous deficits have been linked to MD, including poor number sense (Butterworth, 1999), visuo-spatial difficulties (Rourke and Conway, 1997) and executive dysfunction (Geary et al., 2007a) and as a group the MDRD subtype may have had superior skills to MD in any of these areas.

The possibility that these subtypes were relying on different strategies when completing the different mathematical tasks has previously been suggested (Hanich et al., 2001). While this is a somewhat speculative suggestion, a correlational analysis performed in the present research does lend support to this idea. Phonological ability was consistently highly associated with the performance of MD on each of the mathematical tasks, whereas non-verbal and verbal ability were not. It may seem surprising that phonological ability was related to maths performance much more than verbal ability despite both being language-based tasks. However, compared to the verbal IQ tasks used in the present study, the phonological tasks require very basic skills, for example, rhyming and the ability to break words down into phonemes (Muter et al., 1997). In contrast, the verbal subtests of the British Ability Scales require a broad range of higher order skills such as vocabulary knowledge, reasoning, and abstract thinking (Elliott et al., 1997). By contrast only non-verbal ability predicted the performance of the RD subtype on each of the mathematical tasks. Similarly, non-verbal ability was a better predictor than verbal ability of MDRD children's performance on most tasks. These findings suggest that the children with MD may tend to use their intact verbal skills more often than their impaired non-verbal skills to solve problems. On the other hand, the RD subtype may use their intact non-verbal skills more than their weak verbal skills to solve problems. These findings indicate that language does not play a 'standard' role in mathematical tasks, rather the role of language will vary from individual to individual depending on their particular strengths and weaknesses. Indeed, cross-cultural evidence shows that amongst cultures where counting words are not available, children solve non-verbal calculation problems using spatial strategies. In contrast English-speaking children hardly ever use spatial strategies and tend to rely more on counting words (Butterworth et al., 2011).

Greater knowledge of individual differences in strategy use would allow interventionists to design interventions based on the strength and weaknesses of the child (Dowker and Sigley, 2010) rather than forcing them to use 'standard procedures' which may not suit their learning style. For example, students with specific RD often have difficulty recalling number facts (e.g., Simmons and Singleton, 2006; Smedt and Boets, 2010), and for these students use of derived strategies based on facts that they can recall may be more appropriate. In some cases students will need assistance to develop appropriate strategies and in other cases they may come up with their own strategies. For example, university students with specific RDs mention developing their own visual

strategies (e.g., diagrams) to understand and solve mathematical problems and to compensate for their relatively weak verbal skills (Perkin and Croft, 2007). There has been some research on how children with uneven abilities solve exact calculation compared to TA children (e.g., Geary et al., 2000; Jordan et al., 2003; Wylie et al., 2012). Generally speaking these studies show that children with MD and MDRD employ a different strategy mix to RD or TA when solving problems, either by relying on developmentally immature strategies or trying to use mature counting strategies before developmentally ready. However, less is known about the use of individual strategies on other mathematical tasks (e.g., place value, geometry). In addition, asking children about how they solve problems can only identify different procedures, it does not tell us about individual differences in terms of how children represent number in the brain. While much is now known about the neural basis of numerical cognition (Butterworth and Walsh, 2011), less is known about how children with uneven abilities represent mathematical problems at a neural level compared to TA children.

The performance of TA on each of the tasks was correlated with phonological, verbal and non-verbal ability, to indicate the language and non-verbal requirements of these tasks for children with good verbal and non-verbal skills who are more likely to follow standard procedures. For TA children, the correlation analyses did not highlight any clear bias towards verbal or non-verbal strategy use. In contrast to previous research (Dowker, 1998), verbal ability did not predict the performance of TA children on most mathematical tasks. It could be the case that as children get older and their verbal skills develop further they are better able to utilize these skills when solving mathematics problems. If so, this may partially account for the stronger relationship between maths and verbal IQ observed in Dowker's sample which comprised children aged 5–9 years. It was surprising that for TA verbal and non-verbal ability did not relate more consistently with the mathematical tasks; however, the correlations may have been weakened by ceiling effects on the mathematical tasks.

A key aim of the present research was to evaluate the suitability of subtyping as an approach to examining the role of language in mathematics. On a positive note, subtyping has greater ecological validity than correlational analyses, in the sense that children are arbitrarily classed as having MD in the classroom. Indeed, decisions regarding whether or not to intervene are often made based on these arbitrary cut-off points. However, in contrast to correlational approaches, subtyping does not use full variation in statistical analysis. It is important to note that a key limitation of the present study and the previous work of Hanich et al. (2001) and Jordan et al. (2003), was the use of subtyping classification based on an assessment at a single time point. Research on subtype stability has shown that while some young children have persistent MDs, others have a more variable pattern of achievement and can be mislabeled if assessed only once (Mazzocco and Myers, 2003). It is possible that the lenient cut-off point (35th percentile) used in the present analysis may have affected the results. Indeed, Geary et al. (2007b) found that children with mathematical disabilities (<15th percentile) and those with low maths achievement (23rd–39th percentile) displayed qualitatively different profiles of deficit. However, Jordan and Hanich (2003) found that children with

below average (<15th percentile) and those with low (15th–30th percentile) mathematics achievement displayed qualitatively similar performance on a range of mathematical tasks.

The present analysis has identified a further limitation of using a subtyping approach. Assessing the language requirements of these tasks based on subtyping comparisons is difficult because in the present study, and in Hanich et al.'s (2001) investigation, on some occasions the RD subtype was significantly impaired, yet the MDRD subtype performed at a similar level to the MD subtype. The opposite situation was also observed by Hanich et al. (2001), where the MD subtype significantly outperformed the MDRD subtype yet the RD subtype was not significantly impaired. These inconsistencies indicate that subtyping on its own as a methodology does not give a good indication of the verbal/non-verbal requirements of a task. Indeed, Bartelet et al. (2014) have concluded that it is difficult to draw conclusions from subtyping evidence alone due to the heterogeneous nature of MD. Despite these limitations, subtyping in conjunction with correlational evidence does provide important insights into the role of language in mathematics. The findings from the present study suggest that children can achieve very similar performance levels via different mixes of verbal and non-verbal strategies. Consistent with the existing body of research on mathematical tasks (e.g., Dowker, 2005; Dowker et al., 2008; LeFevre et al., 2010), subtypes with weak verbal or non-verbal ability do not perform as well as their typically achieving counterparts, suggesting that both language and non-verbal skills are important in achieving age-appropriate performance on most tasks.

## ACKNOWLEDGMENT

This research was supported by a grant from the Department of Employment and Learning (NI).

## REFERENCES

- Adams, M. (1990). *Beginning to Read – Thinking and Learning about Print*. Cambridge, MA: The MIT Press.
- Arbuckle, J. (2006). *AMOS (Version 7.0) Computer Software and Manual*. Chicago: Smallwaters.
- Barrouillet, P., Fayol, M., and Lathuliere, E. (1997). Selecting between competitors in multiplication tasks: an explanation of the errors produced by adolescents with learning disabilities. *Int. J. Behav. Dev.* 21, 253–275. doi: 10.1080/016502597384857
- Bartelet, D., Ansari, D., Vaessen, A., and Blomert, L. (2014). Cognitive subtypes of mathematics learning difficulties in primary education. *Res. Dev. Disabil.* 35, 657–670. doi: 10.1016/j.ridd.2013.12.010
- Bollen, K., and Curran, P. (2006). *Latent Curve Models: A Structural Equation Perspective*. Hoboken, NJ: John Wiley and Sons.
- Browne, M., and Cudeck, R. (1993). “Alternative ways of assessing model fit,” in *Testing Structural Equation Models*, eds K. Bollen and S. Long (Newbury Park, NJ: Sage).
- Bruandet, M., Molko, N., Cohen, L., and Dehaene, S. (2004). A cognitive characterization of dyscalculia in Turner syndrome. *Neuropsychologia* 42, 288–298. doi: 10.1016/j.neuropsychologia.2003.08.007
- Butterworth, B. (1999). *The Mathematical Brain*. London: Macmillan.
- Butterworth, B. (2008). Numerosity perception: how many speckles on the hen? *Curr. Biol.* 18, 388–389. doi: 10.1016/j.cub.2008.03.014
- Butterworth, B., Reeve, R., and Reynolds, F. (2011). Using mental representations of space when words are unavailable: studies of enumeration and arithmetic in indigenous Australia. *J. Cross Cult. Psychol.* 42, 630–638. doi: 10.1177/0022022111406020
- Butterworth, B., and Walsh, V. (2011). Neural basis of mathematical cognition. *Curr. Biol.* 21, R618–R621. doi:10.1016/j.cub.2011.07.005
- Catts, H. W., Adlof, S. M., Hogan, T. P., and Weismer, S. E. (2005). Are specific language impairment and dyslexia distinct disorders? *J. Speech Lang. Hear. Res.* 48, 1378–1396. doi: 10.1044/1092-4388(2005)096
- Dehaene, S., and Cohen, L. (1991). Two mental calculation systems: a case study of severe acalculia with preserved approximation. *Neuropsychologia* 29, 1045–1074. doi: 10.1016/0028-3932(91)90076-K
- Dehaene, S., Spelke, E., Pineda, P., Stanescu, R., and Tsivkin, S. (1999). Sources of mathematical thinking: behavioral and brain-imaging evidence. *Science* 284, 970–974. doi: 10.1126/science.284.5416.970
- Donlan, C., Cowan, R., Newton, E. J., and Lloyd, D. (2007). The role of language in mathematical development: evidence from children with specific language impairments. *Cognition* 103, 23–33. doi: 10.1016/j.cognition.2006.02.007
- Dowker, A. (1994). Adults with mild specific calculation difficulties. *Paper presented at the British Psychological Society London Conference*, London.
- Dowker, A. (1998). “Individual differences in arithmetical development,” in *The Development of Mathematical Skills*, ed. C. Donlan (London: Taylor & Francis).
- Dowker, A. (2003). “Young children’s estimates for addition: the zone of partial knowledge and understanding,” in *The Development of Arithmetic Concepts and Skills: Constructing Adaptive Expertise*, eds A Baroody and A Dowker (Mahwah, NJ: Lawrence Erlbaum Associates, Publishers).
- Dowker, A. (2004). *What Works for Children with Mathematical Difficulties?* Oxford: University of Oxford.
- Dowker, A. (2005). *Individual Differences in Arithmetic*. Hove: Psychology Press. doi: 10.4324/9780203324899
- Dowker, A., Bala, S., and Lloyd, D. (2008). Linguistic influences on mathematical development: how important is the transparency of the counting system? *Philos. Psychol.* 21, 523–528. doi: 10.1080/09515080802285511
- Dowker, A., and Sigley, G. (2010). Targeted interventions for children with arithmetical difficulties. *Br. J. Educ. Psychol. Monogr. II* 7, 65–81. doi: 10.1348/97818543370009X12583699332492
- Elliott, C., Smith, P., and McCulloch, K. (1997). *British Ability Scales 2*. London: NFER-Nelson.
- Fazio, B. (1999). Arithmetic calculation, short-term memory, and language performance in children with specific language impairment: a 5-year follow-up. *J. Speech Lang. Hear. Res.* 42, 420–431. doi: 10.1044/jslhr.4202.420
- Fuchs, L., Comptom, D., Fuchs, D., Paulsen, K., Bryant, J., and Hamlet, C. (2005). The prevention, identification, and cognitive determinants of math difficulty. *J. Educ. Psychol.* 97, 493–513. doi: 10.1037/0022-0663.97.3.493
- Fuchs, L., Fuchs, D., Comptom, D., Powell, S., Seethaler, P., Capizzi, A., et al. (2006). The cognitive correlates of third-grade skill in arithmetic, algorithmic computation, and word problems. *J. Educ. Psychol.* 98, 29–43. doi: 10.1037/0022-0663.98.1.29
- Geary, D. (1990). A componential analysis of an early learning deficit in mathematics. *J. Exp. Child Psychol.* 49, 363–383. doi: 10.1016/0022-0965(90)90065-G
- Geary, D. (1993). Mathematical disabilities: cognitive, neuropsychological, and genetic components. *Psychol. Bull.* 114, 345–362. doi: 10.1037/0033-2909.114.2.345
- Geary, D., Brown, S., and Samaranyake, V. (1991). Cognitive addition: a short longitudinal study of strategy choice and speed-of-processing differences in normal and mathematically disabled children. *Dev. Psychol.* 27, 787–797. doi: 10.1037/0012-1649.27.5.787
- Geary, D., Hamson, C., and Hoard, M. (2000). Numerical and arithmetical cognition: a longitudinal study of process and concept deficits in children with learning disability. *J. Exp. Child Psychol.* 77, 236–263. doi: 10.1006/jecp.2000.2561
- Geary, D. C., and Hoard, M. (2001). Numerical and arithmetical deficits in learning-disabled children: relation to dyscalculia and dyslexia. *Aphasiology* 15, 635–647. doi: 10.1080/02687040143000113
- Geary, D., Hoard, M., Byrd-Craven, J., Nugent, L., and Numtee, C. (2007a). Cognitive mechanisms underlying achievement deficits in children with mathematical learning disability. *Child Dev.* 78, 1343–1359. doi: 10.1111/j.1467-8624.2007.01069.x
- Geary, D., Hoard, M., and Hamson, C. (1999). Numerical and arithmetical cognition: patterns of functions and deficits in children at risk for a mathematical disability. *J. Exp. Child Psychol.* 74, 213–239. doi: 10.1006/jecp.1999.2515

- Geary, D., Hoard, M., Nugent, L., and Byrd Craven, J. (2007b). "Strategy use, long-term memory and working memory capacity," in *Why is Math so Hard for Some Children?*, eds D. Berch and M. Mazzocco (Baltimore, MD: Brookes).
- Ginsburg, H., and Baroody, A. (1990). *Test of Early Mathematics Ability 2*. Austin, TX: Pro.Ed.
- Ginsburg, H., and Baroody, A. (2003). *Test of Early Mathematics Ability 3*. Austin, TX: Pro.Ed.
- Gobel, S. M., and Snowling, M. J. (2010). Number-processing skills in adults with dyslexia. *Q. J. Exp. Psychol.* 63, 1361–1373. doi: 10.1080/17470210903359206
- Grauberg, E. (1998). *Elementary Mathematics and Language Difficulties*. London: Whurr Publishers.
- Hanich, L., Jordan, N., Kaplan, D., and Dick, J. (2001). Performance across different areas of mathematical cognition in children with learning difficulties. *J. Educ. Psychol.* 93, 615–626. doi: 10.1037/0022-0663.93.3.615
- Jordan, J. A., Mulhern, G., and Wylie, J. (2009). Individual differences in trajectories of arithmetical development in typically developing 5- to 7-year olds. *J. Exp. Child Psychol.* 103, 455–468. doi: 10.1016/j.jecp.2009.01.011
- Jordan, J. A., Wylie, J., and Mulhern, G. (2010). Phonological awareness and mathematical difficulty: a longitudinal perspective. *Br. J. Dev. Psychol.* 28, 89–107. doi: 10.1348/026151010X485197
- Jordan, N., and Hanich, L. (2000). Mathematical thinking in second-grade children with different types of learning difficulties. *J. Learn. Disabil.* 33, 567–578. doi: 10.1177/002221940003300605
- Jordan, N., and Hanich, L. (2003). Characteristics of children with moderate mathematics deficiencies: a longitudinal perspective. *Learn. Disabil. Res. Pract.* 18, 213–221. doi: 10.1111/1540-5826.00076
- Jordan, N., Hanich, L., and Kaplan, D. (2003). A longitudinal study of mathematical competencies in children with specific mathematics difficulties versus children with co-morbid mathematics and reading difficulties. *Child Dev.* 74, 834–850. doi: 10.1111/1467-8624.00571
- Jordan, N., Levine, S., and Huttenlocher, J. (1995). Calculation abilities in young children with different patterns of cognitive functioning. *J. Learn. Disabil.* 98, 53–64. doi: 10.1177/002221949502800109
- Jordan, N., and Montani, T. (1997). Cognitive arithmetic and problem solving: a comparison of children with specific and general mathematics difficulties. *J. Learn. Disabil.* 30, 624–634. doi: 10.1177/00222194970300606
- Kamhi, A. G., and Catts, H. W. (1986). Toward an understanding of developmental language and reading disorders. *J. Speech Hear. Disord.* 51, 337–347. doi: 10.1044/jshd.5104.337
- Koontz, K., and Berch, D. (1996). Identifying simple numerical stimuli: processing inefficiencies exhibited by arithmetic learning disabled children. *Math. Cogn.* 2, 1–23. doi: 10.1080/135467996387525
- Koponen, T., Mononen, R., Räsänen, P., and Ahonen, T. (2006). Basic numeracy in children with specific language impairment: heterogeneity and connections to language. *J. Speech Lang. Hear. Res.* 49, 58–73. doi: 10.1044/1092-4388(2006/005)
- LeFevre, J.-A., Fast, L., Skwarchuk, S.-L., Smith-Chant, B. L., Bisanz, J., and Kamawar, D., et al. (2010). Pathways to mathematics: longitudinal predictors of performance. *Child Dev.* 81, 1753–1767. doi: 10.1111/j.1467-8624.2010.01508.x
- Mazzocco, M. M., and Myers, G. F. (2003). Complexities in identifying and defining mathematics learning disability in the primary school-age years. *Ann. Dyslexia* 53, 218–253. doi: 10.1007/s11881-003-0011-7
- Molko, N., Cachia, A., Riviere, D., Mangrin, J., Bruandet, M., Le Bihan, D., et al. (2003). Functional and structural alterations of the intraparietal sulcus in a developmental dyscalculia of genetic origin. *Neuron* 40, 847–858. doi: 10.1016/S0896-6273(03)00670-6
- Muter, V., Hulme, C., and Snowling, M. (1997). *Phonological Abilities Test*. London: Hodder Education.
- Northern Ireland Statistics and Research Agency. (2005). *Northern Ireland Multiple Deprivation Measure, 2005*. Belfast: NISRA.
- Ostad, S. (1997). Developmental differences in addition strategies: a comparison of mathematically disabled and mathematically normal children. *Br. J. Educ. Psychol.* 67, 345–357. doi: 10.1111/j.2044-8279.1997.tb01249.x
- Ostad, S. (1998). Developmental differences in solving simple arithmetic word problems and simple number-fact problems: a comparison of mathematically normal and mathematically disabled children. *Math. Cogn.* 4, 1–19. doi: 10.1080/135467998387389
- Ostad, S. (1999). Developmental progression of subtraction strategies: a comparison of mathematically normal and mathematically disabled children. *Eur. J. Spec. Needs Educ.* 14, 21–36. doi: 10.1080/0885625990140103
- Ostad, S. (2000). Cognitive subtraction in a developmental perspective: accuracy, speed-of-processing and strategy-use differences in normal and mathematically disabled children. *FOCUS Learn. Prob. Math.* 22, 18–31.
- Perkin, G., and Croft, T. (2007). The dyslexic student and mathematics in higher education. *Dyslexia* 13, 193–210. doi: 10.1002/dys.344
- Pica, P., Lemer, C., Izard, V., and Dehaene, S. (2004). Exact and approximate arithmetic in an Amazonian Indigene group. *Science* 306, 499–503. doi: 10.1126/science.1102085
- Robinson, C., Menchetti, B., and Torgesen, J. (2002). Toward a two-factor theory of one type of mathematics disabilities. *Learn. Disabil. Res. Pract.* 17, 81–89. doi: 10.1111/1540-5826.00035
- Rourke, B., and Conway, J. (1997). Disabilities of arithmetic and mathematical reasoning: perspectives from neurology and neuropsychology. *J. Learn. Disabil.* 30, 34–46. doi: 10.1177/002221949703000103
- Rovet, J., Szekely, C., and Hockenberry, M. (1994). Specific arithmetical calculation deficits in children with Turner syndrome. *J. Clin. Exp. Neuropsychol.* 16, 820–839. doi: 10.1080/01688639408402696
- Russell, R., and Ginsburg, H. (1984). Cognitive analysis of children's mathematical difficulties. *Cogn. Instr.* 1, 217–244. doi: 10.1207/s1532690xci0102\_3
- Siegler, R. (1996). *Emerging Minds: The Process of Change in Children's Thinking*. New York: Oxford University Press.
- Siegler, R., and Shrager, J. (1984). "Strategy choice in addition and subtraction: how do children know what to do?," in *Origins of Cognitive Skills*, ed. C. Sophian (Hillsdale, NJ: Erlbaum).
- Simmons, F. R., and Singleton, C. (2006). The mental and written arithmetic abilities of adults with dyslexia. *Dyslexia* 12, 96–114. doi: 10.1002/dys.312
- Smedt, B., and Boets, B. (2010). Phonological processing and arithmetic fact retrieval: evidence from developmental dyslexia. *Neuropsychologia* 48, 3973–3981. doi: 10.1016/j.neuropsychologia.2010.10.018
- Swanson, H. (2007). "Cognitive aspects of math disabilities," in *Why is Math so Hard for Some Children?*, eds D. Berch and M. Mazzocco (Baltimore, MD: Brookes).
- Swanson, H., and Beebe-Frankenberger, M. (2004). The relationship between working memory and mathematical problem solving in children at risk and not at risk for serious math difficulties. *J. Educ. Psychol.* 96, 471–491. doi: 10.1037/0022-0663.96.3.471
- Venneri, A., Cornoldi, C., and Garuti, M. (2003). Arithmetic difficulties in children with visuospatial learning disability (VLD). *Child Neuropsychol.* 9, 175–183. doi: 10.1076/chin.9.3.175.16454
- Warrington, E. (1982). The fractionation of arithmetical skills: a single case study. *Q. J. Exp. Psychol.* 34, 31–51. doi: 10.1080/14640748208400856
- Woodcock, R., McGrew, K., and Mather, N. (2001). *Woodcock-Johnson III Tests of Achievement*. Rolling Meadows, IL: Riverside Publishing.
- Wylie, J., Jordan, J. A., and Mulhern, G. (2012). Strategic development in exact calculation: group and individual differences in four achievement subtypes. *J. Exp. Child Psychol.* 113, 112–130. doi: 10.1016/j.jecp.2012.05.005
- Ziegler, J. C., Bertrand, D., Toth, D., Csepe, V., Reis, A., Faisca, L., et al. (2010). Orthographic depth and its impact on universal predictors of reading: a cross language investigation. *Psychol. Sci.* 21, 551–559. doi: 10.1177/0956797610363406

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 07 November 2014; accepted: 12 February 2015; published online: 05 March 2015.

Citation: Jordan JA, Wylie J and Mulhern G (2015) Mathematics and reading difficulty subtypes: minor phonological influences on mathematics for 5–7-years-old. *Front. Psychol.* 6:221. doi: 10.3389/fpsyg.2015.00221

This article was submitted to *Developmental Psychology*, a section of the journal *Frontiers in Psychology*.

Copyright © 2015 Jordan, Wylie and Mulhern. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Number processing and arithmetic skills in children with cochlear implants

Silvia Pixner<sup>1\*</sup>, Martin Leyrer<sup>2,3</sup> and Korbinian Moeller<sup>4,5</sup>

<sup>1</sup> Institute of Applied Psychology, UMIT – The Health and Life Sciences University, Hall in Tyrol, Austria

<sup>2</sup> Department of Otolaryngology, Paracelsus University Medical School Salzburg, Salzburg, Austria

<sup>3</sup> Department of Linguistics, University of Salzburg, Salzburg, Austria

<sup>4</sup> Knowledge Media Research Center, Tübingen, Germany

<sup>5</sup> Department of Psychology, University of Tübingen, Tübingen, Germany

## Edited by:

Ann Dowker, University of Oxford, UK

## Reviewed by:

Teresa Mitchell, University of Massachusetts Medical School, USA  
Chris Donlan, University College London, UK

## \*Correspondence:

Silvia Pixner, Institute of Applied Psychology, UMIT – The Health and Life Sciences University, Eduard Wallnoefer Zentrum 1, 6060 Hall in Tyrol, Austria  
e-mail: [silvia.pixner@umit.at](mailto:silvia.pixner@umit.at)

Though previous findings report that hearing impaired children exhibit impaired language and arithmetic skills, our current understanding of how hearing and the associated language impairments may influence the development of arithmetic skills is still limited. In the current study numerical/arithmetic performance of 45 children with a cochlea implant were compared to that of controls matched for hearing age, intelligence and sex. Our main results were twofold disclosing that children with CI show general as well as specific numerical/arithmetic impairments. On the one hand, we found an increased percentage of children with CI with an indication of dyscalculia symptoms, a general slowing in multiplication and subtraction as well as less accurate number line estimations. On the other hand, however, children with CI exhibited very circumscribed difficulties associated with place-value processing. Performance declined specifically when subtraction required a borrow procedure and number line estimation required the integration of units, tens, and hundreds instead of only units and tens. Thus, it seems that despite initially atypical language development, children with CI are able to acquire arithmetic skills in a qualitatively similar fashion as their normal hearing peers. Nonetheless, when demands on place-value understanding, which has only recently been proposed to be language mediated, hearing impaired children experience specific difficulties.

**Keywords:** number processing, multiplication, number line estimation, subtraction, cochlear implants

## INTRODUCTION

At first glance, skills like mental arithmetic and/or magnitude comparison are not readily dependent on language abilities. In fact, quite many numerical competencies – such as numerical discrimination and additive/subtractive expectations – are mastered well before children are able to actively produce their first words (for a respective review, see Feigenson et al., 2004). Thus, the lack of studies investigating numerical cognition in individuals with impaired hearing (who are known to have a delayed or even aberrant language development) comes to no surprise. The present study aims to address the case of deaf and profoundly hearing impaired children who received a cochlear implant (CI) during early infancy.

Recent research shows unambiguously that early implantation enables most of the young users with CI to develop considerable speech and language competence (e.g., Geers et al., 2003; Nott et al., 2003; Tomblin et al., 2005; Connor et al., 2006; Johnson and Goswami, 2010). The developmental trajectories, however, seem to be deviant from typical with respect to phonetics and phonology and delayed with respect to grammar and lexicon (e.g., Blarney et al., 2001; Chin, 2006; Leyrer, 2008; Adi-Bensaid and Tubul-Lacy, 2009; Geers et al., 2009; Friedmann and Szterman, 2011). Importantly, such variability in developmental pathways of linguistic skills might affect other cognitive domains such as numerical cognition. In the present study, we will evaluate the

benefits and limitations of cochlear implantation for the acquisition of numerical skills in affected children. Before presenting the experimental study the interrelation of numerical and language skills will be elaborated on briefly.

## THE INTERRELATION OF NUMERICAL AND LANGUAGE SKILLS

It has been argued that language plays a key role in the development of number-related language processing and in particular so in the development of number concepts (e.g., Carey, 2004). Following this, LeFevre et al. (2010) proposed a developmental calculation model that differentiates three relevant pathways for the development of numerical skills: a linguistic pathway, a quantitative and a spatial attention pathway. In their sample of 182 children aged 4.5–7.5 years the three aforementioned pathways were found to contribute independently to early numerical skills. Moreover, there is also evidence suggesting influences of language skills to be rather specific. The currently most influential model of number processing – the Triple Code model – (see Dehaene et al., 2003; Arsalidou and Taylor, 2011 for latest amendments) – suggests numerical information to be represented by three codes within the human brain. The visual-Arabic number form is the most basic code. It is recruited to perceive digits as numerically informative symbols and associated with bilateral occipital brain areas. Additionally, the Triple Code model differentiates between an analogue quantity code and verbal numerical

representations. The analogue quantity code is involved whenever quantity or magnitude information of numbers is processed. It is assumed to be subserved by bilateral cortex areas around the intra-parietal sulcus. Finally, verbal numerical representations are recruited in tasks such as number naming. Additionally, arithmetic facts (e.g., multiplication fact knowledge) are assumed to be stored in a verbal code. Verbal numerical representations are associated with left-lateralized perisylvian language areas and the angular gyrus. In line with the assumption of a verbal numerical representation language influences should be most prominent when it comes to arithmetic fact knowledge, whereas representations of numerical quantity and numerical symbols should be less dependent on language. Noteworthy, the findings of Koponen et al. (2006) revealed that number naming speed is indeed closely related to arithmetic fact retrieval. This means that skilled calculators are able to directly retrieve the result of a number fact (e.g.,  $3 \times 4$ ) from phonological long-term memory without having to apply procedural calculation strategies. Additionally, it has been observed that this fact retrieval processes are subject to interference by concurrent articulation (Lee and Kang, 2002; Moeller et al., 2011). This provides further evidence for the verbal (language related) format of arithmetic fact knowledge. Nevertheless, also more direct evidence for the close interrelation between number fact retrieval and language abilities is accumulating. For instance, Fazio (1999) reported deficient fact retrieval skills in children diagnosed with a specific language impairment (SLI, see also Donlan et al., 2007). Furthermore, it has been found repeatedly that children suffering from dyslexia (whose core difficulty by definition is impaired acquisition of written language) often exhibit deficient number fact retrieval, too (e.g., Snowling, 2000; see also Miles et al., 2001; Simmons and Singleton, 2008; De Smedt and Boets, 2010; Göbel and Snowling, 2010). Finally, beyond the case of fact retrieval, counting abilities have also been observed to be associated with language competencies (children with dyslexia: Simmons and Singleton, 2008; children with SLI: Koponen et al., 2006; Donlan et al., 2007).

Nevertheless, as already indicated in the Triple Code model not all aspects of numerical cognition should be associated with language competencies. In line with this suggestion it was found that subitizing (LeFevre et al., 2010), symbolic calculation (McNeil and Burgess, 2002), number comparison (O'Hearn and Luna, 2009) as well as number line estimation (requiring children to estimate the position of a given number on a presented number line; Koponen et al., 2006) are rather independent from language skills. However, even when these tasks are referred to as being non-verbal, recent research indicated that this might only be part of the story. For instance Pixner et al. (2011a) disclosed language influences on a two-digit number comparison task that was administered to German, Italian and Czech-speaking first graders. The three groups are ideal populations to study language influences because they are distinguishable regarding the correspondence between symbolic Arabic and spoken number word systems. While spoken and the symbolic Arabic number systems closely correspond to each other in Italian (*venti-cinque*/twenty-five  $\rightarrow$  25], the German number word system is intransparent insofar as the order of tens and units is inverted in spoken as compared to symbolic notation

(*fünfundzwanzig*/five-and-twenty  $\rightarrow$  25). Finally, in Czech both inverted and non-inverted number words are utilized. Noteworthy, the intransparency of the inverted number word system posed particular difficulty on German-speaking children's transcoding performance in general (Zuber et al., 2009) and Czech-speaking children when asked to transcode inverted number words (see also Pixner et al., 2011a for language effects in magnitude comparison). In line with this, mental number line representations also seem to be moderated by language characteristics (Helmreich et al., 2011). LeFevre et al. (2010) propose in their model of numerical development that the number line task calls on both semantic (number magnitude) and spatial representations. Additionally, the authors suggest that performing the number line estimation task for multi-digit number ranges requires mastery of the base-10 structure of the Arabic number system (see also Moeller et al., 2009) which in turn is clearly language dependent as indicated by recent evidence from transcoding (Zuber et al., 2009; Pixner et al., 2011b) but also number line estimation (Helmreich et al., 2011). Taken together, though there is accumulating evidence for a link between language and number processing knowledge on the exact nature and the underlying mechanisms of this association is still rather patchy even in typically developing not to say in atypically developing children such as deaf or children with CI.

#### NUMBER PROCESSING IN HEARING IMPAIRED AND DEAF INDIVIDUALS

A frequent observation in educational settings is that children with profound hearing impairments as well as deaf children quite often experience difficulties to acquire calculation skills (e.g., Zarfaty et al., 2004; Ansell and Pagliaro, 2006). Upon taking into account the aforementioned link between language and numerical skills (typically developing children: Zuber et al., 2009; Helmreich et al., 2011; Pixner et al., 2011a,b; children with dyslexia: Simmons and Singleton, 2008; children with SLI: Koponen et al., 2006; Donlan et al., 2007) number-related deficiencies in hearing impaired individuals come to no surprise. Interestingly, already the National Council of Teachers of the Deaf Research Committee, 1957 examined 200 deaf students in Great Britain and found a significant (i.e., a 1–21/2 year) delay in the acquisition of arithmetical skills. Similar findings are reported by Kramer (2007) who investigated German-speaking deaf individuals. However, different from research on typically developing children for which the interrelation of language and number processing has been investigated quite specifically, there is a scarcity of studies systematically investigating specific numerical skills (e.g., arithmetic fact retrieval, basic arithmetic, number line estimation, etc.) and its relation to language skills for hearing impaired individuals in a comparative manner. Therefore, the present study pursued this issue.

#### THE PRESENT STUDY

The main aim of the present study was to systematically examine numerical and arithmetical skills in formerly deaf children that received a CI in early childhood as compared to typically developing children. It is important to note that there is broad consensus in the literature indicating atypical language development in children with CI (e.g., Boothroyd et al., 1991; Dawson et al., 1995; Geers et al., 2009; Pisoni et al., 2010; Ingvalson and Wong, 2013;

Rinaldi et al., 2013). Therefore, we did not wish to evaluate the influence of an atypical language development in children with CI on their numerical/arithmetic abilities by correlating their performance in a language task with those in numerical/arithmetic tasks. Instead, we compared the performance of children with CI and a control group matched on hearing age and general intellectual functioning in specific numerical/arithmetic tasks chosen for either their strong dependence on language-related processing (i.e., multiplication fact retrieval) or their only weak dependence on language skills (i.e., two-digit subtraction and number line estimation). Thereby, we aimed at evaluating whether impairments possibly observed for CI children may be associated with their known atypical language development. In particular, the following research questions and hypotheses were pursued.

- (1) On a general level we were interested in whether the number of students with poor arithmetic skills is comparable for children with CI and their normal hearing (NH) peers. Following the rationale that numerical competencies are directly (arithmetic facts) or indirectly (magnitude representation) related to verbal language representations it is expected that children with CI are more likely to score poor on a standardized mathematics performance test as compared to their hearing peers.
- (2) More specifically, it was of interest whether possible impairments of children with CI would be most pronounced for those tasks for which the closest association with language is assumed. When this is the case, children with CI should perform significantly worse than their hearing peers in a multiplication task assessing verbally driven arithmetic fact retrieval. On the other hand, for tasks with a more indirect association with language processes such as subtraction and/or number line estimation our hypotheses need to be more specific. Generally, children with CI should not perform worse on subtraction. However, this might be moderated by the need for a borrowing. Following the model of LeFevre et al. (2010) mastery of the place-value structure of the Arabic number system is language dependent (see also Pixner et al., 2011a,b). Thus, as successful application of the borrowing procedure requires place-value understanding (see Klein et al., 2010a,b for the case of carry in addition) we expect children with CI to experience particular difficulties for borrowing subtraction problems. Similarly, LeFevre et al. (2010) suggest performance in the number line estimation task to be associated with place-value understanding (see also Moeller et al., 2009; Helmreich et al., 2011). Thus, children with CI should also show poorer performance in the number line estimation task as compared to their NH peers.

## MATERIALS AND METHODS

### PARTICIPANTS

Overall, 94 children participated in the present study, 45 children (26 males) with CI and 49 NH controls (23 males). Cochlear implants were surgically placed when children were at a minimum of eight and a maximum of 50 months of age. Participating children attended third to fifth grade. Fourteen of the hearing impaired participants were in special educational schools. In these schools children with CI are taught together with deaf children.

Even though it is tried to teach children with CI using spoken language as far as possible part of the instruction also includes sign language.

Though children with CI were older than their hearing peers, the two groups were comparable with respect to hearing experience and grade level (see **Table 1**). Please note that we used hearing age instead of age at implantation as a variable describing the hearing/language experience of the children with CI because it is possible to match the two groups on this variable. As we aimed at matching our two groups as closely as possible we decided to use hearing age because it is not possible to match the two groups on age at implantation as a measure of hearing/language experience.

Moreover, in Austria hearing impaired children are likely to start school with a delay of one to 2 years. Thus, grade level was used as matching criteria between the experimental (CI) and the control group (NH) because grade level is more relevant (in terms of the received mathematical instruction) than chronological age when investigating arithmetical skills. Furthermore, the two study groups were comparable with respect to overall intellectual functioning and verbal working memory. For central executive (CE) functioning there even was an advantage for the Children with CI (see **Table 1**).

### ASSESSMENT AND PROCEDURE

The study was approved by the local ethics committee of the UMIT, Hall in Tyrol.

To assess *math achievement* children with CI had to complete the basic arithmetic operations scale of the Heidelberger Rechen-test 1–4 (HRT 1–4; Haffner et al., 2005). Please note that control children were not administered the HRT.

Instead, both groups were asked to solve two PC-administered tasks tapping multiplication and subtraction skills. In addition, a number line estimation task was presented in paper-pencil format. Each child was tested individually in a separate room.

*Multiplication* capabilities were assessed by a verification task comprising 80 multiplication problems with one-digit operands. These critical trials were presented in randomized order preceded by 10 practice trials to ensure task comprehension. Stimuli were presented centrally on the screen in the form  $x \times y = z$  (Arial font, size 48). On each trial, the problem was presented simultaneously with an either correct or incorrect solution probe. Additionally, incorrect probes were separated into two error types: operand errors representing the correct solution to a neighboring multiplication problem of the same table (e.g.,  $3 \times 4 = 15$ ) and so-called non-table errors not related to any item of the multiplication tables (e.g.,  $3 \times 4 = 13$ ). Children were asked to indicate by button press whether the solution probe was correct (right-hand button press) or not (left-hand button press). Stimulus presentation was preceded by a fixation cross presented at the center of the screen for 500 ms. Then the multiplication problem appeared and stayed visible until one of the response buttons was pressed. After an inter-stimulus-interval of 500 ms the fixation cross for the next trial was presented.

*Subtraction* skills were assessed by a choice reaction task involving 40 subtraction problems. Comparable to the multiplication task critical trials were presented in randomized order and were

**Table 1 | Demographic variables and background information on the study groups (mean  $\pm$  SD).**

	Children with CI <i>n</i> = 45	Control group <i>n</i> = 49	Statistical difference
Age (months)	122.18 $\pm$ 11.58	101.86 $\pm$ 16.29	$t(92) = 6.94; p < 0.001$
Hearing experience (months)	98.78 $\pm$ 15.35	101.86 $\pm$ 16.29	$t(92) = 1.55; n.s.$
Grade placement	Third grade <i>n</i> = 17 Fourth grade <i>n</i> = 21 Fifth grade <i>n</i> = 7	Third grade <i>n</i> = 20 Fourth grade <i>n</i> = 29	$t(92) = 1.49; n.s.$
Intellectual functioning (T-score) <sup>°</sup>	53.00 $\pm$ 9.28 <sup>°</sup>	52.69 $\pm$ 4.81 <sup>§</sup>	$t(92) < 1; n.s.$
Verbal working memory (forward span) <sup>#</sup>	4.93 $\pm$ 0.91	4.71 $\pm$ 0.68	$t(92) = 1.33; n.s.$
Central executive (CE) functioning (backward span) <sup>#</sup>	3.80 $\pm$ 0.89	2.92 $\pm$ 0.67	$t(92) = 5.43; p < 0.001$

<sup>°</sup> Measured by CFT 20-R (Weiß, 2008).

<sup>§</sup> Measured when children attended first grade by the CFT-1 (Cattell et al., 1997). As the CFT is thought to tap the *g*-factor of intelligence (that is considered to be rather stable across development) we are confident that intellectual functions are comparable between participant groups.

<sup>#</sup> Measured by a letter span task.

preceded by 10 practice trials. Importantly, in order to assess procedural solution strategies rather than direct fact retrieval (as dominant in multiplication) the subtraction task comprised two-digit numbers only. Problems were presented in the form  $xx - xx$  (Arial font, size: 48) at the  $x/y$ -coordinates (512/300) with the two solution probes appearing below the problem, either on the left side ( $x/y$  coordinates 300/550) or on the right side ( $x/y$  coordinates 724/550). Children had to single out the correct solution by button press (right or left button press). After a fixation cross was presented for 500 ms the problem and the solution probes appeared on the screen simultaneously. The stimuli stayed visible until one of the response buttons was pressed, directly followed by the fixation cross of the next trial. Subtractions were categorized into those requiring a borrow procedure (e.g., 52–37) and those not requiring a borrow procedure (e.g., 49–34). Importantly, problem size was matched between these item categories.

*Number line estimation* performance was assessed in a paper-pencil version of the number-to-position number line estimation task. Children were asked to estimate the spatial position of a given number on a number line ranging from 0 to 100 for a first set of items and from 0 to 1000 for a second set of items (each line measuring 10 cm). Only the start (0) and end (100 or 1000) point of the number line was specified by the respective Arabic number. Above each number line the target number was written in Arabic notation. Overall, 36 critical trials were presented ( $n = 18$  per range) that were preceded by two practice items.

## SCORING AND ANALYSES

### HRT

Performance in the standardized calculation test was scored according to the procedure described in the test manual (converting raw scores into T-scores).

### Multiplication

Multiplication performance was analyzed in two separate two-way ANOVAs. In a first ANOVA the within-subject factor problem

type (correct vs. incorrect) and the between subject factor participant group (CI vs. NH) were discerned. In the second ANOVA influences of the within-subject factor error type (operand vs. non-table error) and the between-subject factor participant group were evaluated. As the two participant groups differed reliably with respect to CE functioning this variable was incorporated as a covariate. Both analyses were conducted separately for reaction times (RT) and error rates.

### Subtraction

Subtraction performance was evaluated using a two-way ANOVA incorporating the within-subject factor task borrowing (with vs. without) and the between-subject factor participant group (CI vs. NH). As for multiplication performance, this analysis was run for RT and error rates and CE functioning was considered as a covariate.

### Number line estimation

Data analysis of the number line task considered the absolute estimation error (i.e., how far the actually indicated position deviated from the correct position of a target number). As two different number lines (i.e., ranging from 0 to 100 and from 0 to 1000) were employed in the current study and in order to make results comparable, the percent absolute estimation error [PAE; i.e., (target number – estimated number)/number range] per number line range was used for further analyses (cf. Siegler and Booth, 2004). Finally, number line estimation performance in terms of PAE was evaluated in a two-way ANOVA with the within-subject factor number line range (0–100 vs. 0–1000) and the between-subject factor participant group (CI vs. NH). Again, CE functioning was incorporated as a covariate.

## RESULTS

### HRT

Seven out of 45 children with CI exhibited considerably poor performance on the index scale *basic arithmetic operations* (as indicated by a percentile <10). Please note that a percentile <10 is

used as cut-off for the diagnostic criteria of developmental dyscalculia (International Classification of Diseases/ICD 10: Dilling and Freyberger, 2001; Diagnostic and Statistical Manual of Mental Disorders/DSM IV: American Psychiatric Association [APA], 1994). Hence, seven children of our experimental group (i.e., 15.6%) would fall below the diagnostic threshold of developmental dyscalculia. Interestingly, another seven children of the experimental group were found to exhibit excellent performance levels on the standardized calculation test (as indicated by percentiles >85). Nevertheless, a statistical evaluation indicated that the T-scores on the HRT index scale basic arithmetic operations of the experimental group ( $M = 48.40$   $SD = 10.49$ ) did not differ reliably from those of the standardization sample of the HRT [ $M = 50.00$ ,  $SD = 10.00$ ;  $t(44) = 0.89$ , *n.s.*].

## MULTIPLICATION

### Analyses of problem type

Analyses of error rates did not reveal any significant result. Neither the main effects of problem type or participant group nor their interaction turned out to be statistically reliable (all  $F < 1$ ). There was also no significant influence of the covariate ( $F < 1$ ).

With respect to response latencies the ANOVA revealed a significant main effect of problem type [ $F(1,91) = 14.61$ ,  $p < 0.001$ ]. This indicated faster responses for accepting a correct solution probe than rejecting an incorrect one (2773 ms vs. 3280 ms, respectively). Additionally, the reliable main effect of participant group [ $F(1,91) = 9.45$ ,  $p < 0.01$ ] indicated that latencies of children with CI were longer than those of NH controls (3516 ms vs. 2537 ms, respectively, see **Figure 1**). The interaction of problem type and group was not significant ( $F < 1$ ). Finally, the influence of the covariate was significant [ $F(1,91) = 8.41$ ,  $p < 0.01$ ] with shorter latencies being associated with higher CE scores.

### Analyses of error type

Again, the ANOVA on error rates did not reveal any significant result. Neither the main effects of error type or participant group

nor their interaction turned out to be statistically reliable (all  $F < 1.25$ , all  $p > 0.27$ ). Also the influence of the covariate was not reliable ( $F < 1$ ).

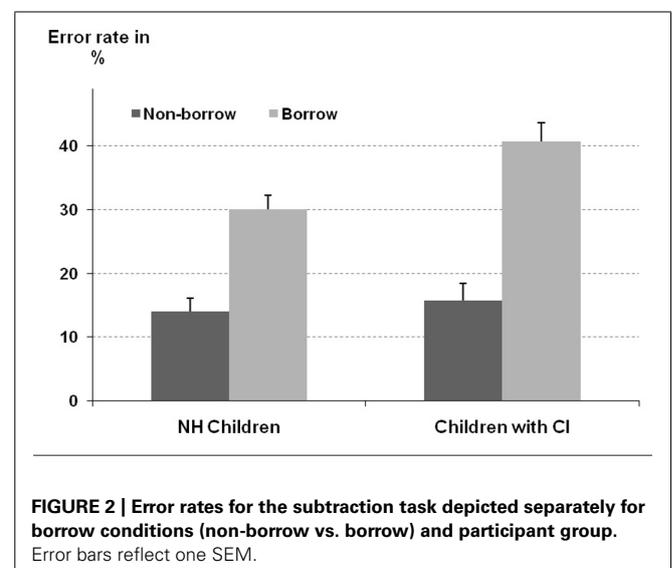
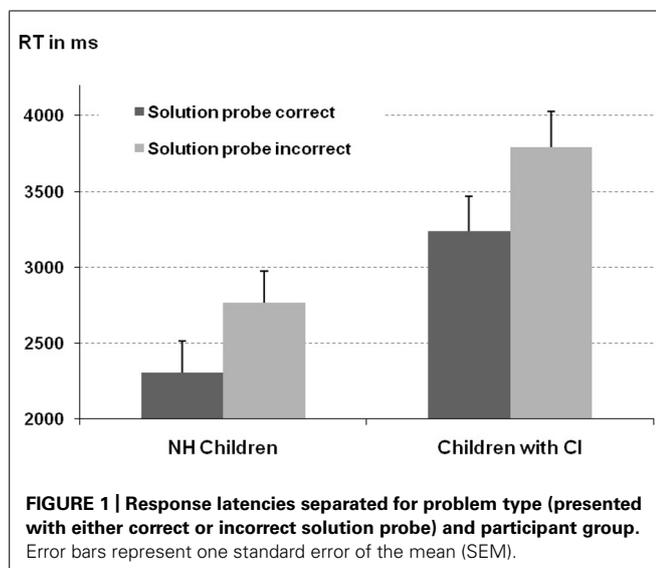
The analysis of participants' response latencies revealed a significant main effect of participant group [ $F(1,91) = 8.41$ ,  $p < 0.01$ ] indicating longer latencies for children with CI than for hearing controls (3794 ms vs. 2767 ms, respectively). Neither the main effect of error type nor the interaction of group and error type was statistically reliable (both  $F < 1$ ). However, the influence of the covariate was significant [ $F(1,91) = 8.70$ ,  $p < 0.01$ ] associating higher CE scores with faster responses.

Taken together this indicates that responses of children with CI were reliably delayed. However, both groups exhibited comparable performance profiles (regarding problem type and error type).

## SUBTRACTION

For error rates the ANOVA revealed a significant borrow effect [ $F(1,91) = 23.86$ ,  $p < 0.001$ ]. As depicted in **Figure 2**, children committed significantly more errors on problems requiring a borrow than on problems not requiring a borrow procedure (35.4% vs. 14.8% errors, respectively). Moreover, the main effect of participant group was not reliable [ $F(1,91) = 2.82$ ,  $p = 0.10$ ] indicating that children with CI did not exhibit a significantly higher error rate than their NH peers (28.2% vs. 22.0% errors, respectively). Most importantly and in line with our expectations the significant interaction of borrow and participant group [ $F(1,91) = 4.82$ ,  $p < 0.05$ ] indicated that the borrowing effect was indeed more pronounced for children with CI as compared to NH controls (25.0% vs. 16.1% errors, respectively, see **Figure 2**). Finally, the influence of the covariate was reliable [ $F(1,91) = 6.03$ ,  $p < 0.05$ ] with a higher CE score being associated with a smaller error rate.

For response latencies only the main effect of participant group was reliable [ $F(1,91) = 12.82$ ,  $p < 0.01$ ]. This indicated that responses of children with CI responded were generally slower as compared to the responses of the control group (6691 ms vs. 5265 ms, respectively). Neither the main effect of borrowing



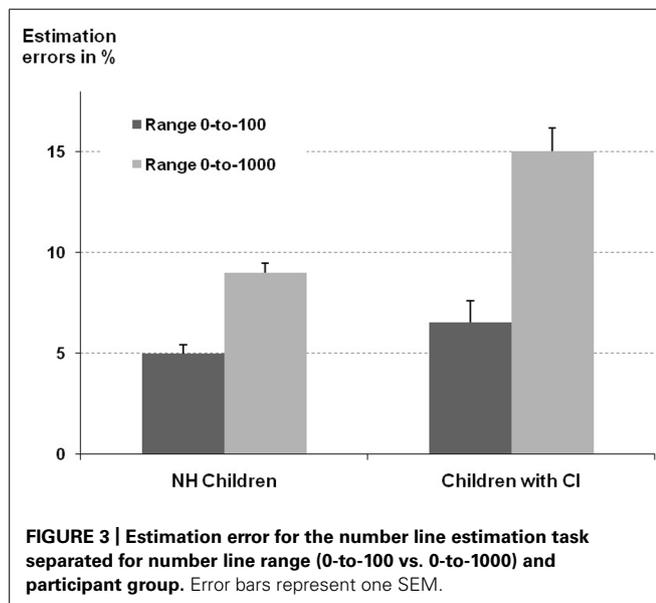
nor the interaction of borrowing and participant group turned out to be reliable (both  $F < 1.26$ , both  $p > 0.26$ ). Additionally, the influence of the covariate was not reliable [ $F(1,91) = 2.48$ ,  $p = 0.12$ ].

In summary, children with CI not only exhibited prolonged response latencies but also experienced difficulties when it comes to the specific processing of place-value information as required by subtraction problems incorporating a borrow procedure.

### NUMBER LINE ESTIMATION

With respect to estimation errors the ANOVA revealed a significant main effect of number line range [ $F(1,91) = 32.30$ ,  $p < 0.001$ ]. This indicated that children's estimation error was reliably larger when asked to mark the position of a given number on a number line ranging from 0 to 1000 as compared with those ranging from 0 to 100 (5.7% vs. 12.0% misplacement, respectively). Additionally, the main effect of participant group was also significant [ $F(1,91) = 11.76$ ,  $p < 0.01$ ]: compared to controls the estimation error of children with CI was significantly larger (10.8% vs. 7.0% misplacement, respectively). Moreover, these main effects were qualified by the reliable interaction of number line range and group [ $F(1,91) = 11.77$ ,  $p < 0.01$ ]. The interaction indicated that the increase of estimation error from the 0 to 100 to the 0 to 1000 number line range was more pronounced for children with CI than for hearing controls (8.5% vs. 4.2% increase in misplacement, respectively, see **Figure 3**). Finally, the influence of the covariate was significant [ $F(1,91) = 8.09$ ,  $p < 0.01$ ] indicating that higher CE scores were associated with a smaller estimation error, this means more precise localization of numbers on the number lines.

Summarizing the results for the number line estimation task it has to be noted that as for the subtraction task we observed specific impairments for children with CI as the demands on processing place-value information increased (i.e., from two- to three-digit numbers).



### DISCUSSION

The main aim of the present study was to investigate numerical/arithmetical capabilities in children with CI and to contrast their performance to NH peers of matched hearing age. We were interested in general as well as in specific performance differences between these participant groups. On the general level we expected children with CI to perform reliably worse than their NH peers, as arithmetical capabilities have been shown to be related to language representations and their processing both directly (e.g., arithmetic facts) and/or indirectly (e.g., magnitude representation). Additionally, we hypothesized that children with CI should be specifically impaired on arithmetical competencies with a specific reference to place-value understanding. According to LeFevre et al. (2010) the representation of the place-value structure of the Arabic number system is closely related to language representations (see also von Aster and Shalev, 2007 for a similar view) even in tasks usually considered not to address language-based numerical representations such as subtraction or number line estimation. Generally, the current data corroborated both of our hypotheses. We observed general (i.e., increased rates of dyscalculia indications, prolonged overall RT) as well as specific impairments (i.e., more pronounced borrowing effect) for children with CI. These will be discussed in turn in the following.

### GENERAL NUMERICAL IMPAIRMENTS OF CHILDREN WITH CI

In a first step, arithmetical skills of children with CI were examined by administering a standardized calculation test. Results revealed that 15.6% of children with CI exhibited performance levels being indicative of developmental dyscalculia. With a general prevalence rate of dyscalculia in the general school population of 4–7% (e.g., Badian, 1993; Gross-Tsur et al., 1996; von Aster et al., 2007) indications of dyscalculia were more prominent in children with CI than was to be expected. This increased rate of dyscalculia indications may be interpreted to index a general impairment of arithmetical capabilities in children with CI. Importantly, the overall result pattern for the more specific arithmetical assessment (i.e., multiplication, subtraction, and number line estimation) corroborated such a conclusion. For each of these tasks we observed reliable group differences indicating that children with CI performed more poorly than the hearing controls: children with CI took longer to complete the multiplication verification as well as the subtraction choice reaction task. Moreover, their number line estimations were less accurate as compared to those of the hearing controls. This indicated that children with CI seemed to suffer from a general impairment, in particular, a general slowing combined with reduced accuracy of mental number line representations, as regards their arithmetical capabilities. Taken together, this suggests that the known atypical language development of children with CI (e.g., Boothroyd et al., 1991; Dawson et al., 1995; Geers et al., 2009; Pisoni et al., 2010; Ingvalson and Wong, 2013; Rinaldi et al., 2013) seemed to exhibit a reliable negative influence on their numerical development as previously proposed (e.g., Fazio, 1999; Koponen et al., 2006). This is in line with recent evidence suggesting language to influence numerical tasks as basic as magnitude comparison and number line estimation but also more complex arithmetic in children (e.g., Helmreich

et al., 2011; Pixner et al., 2011a,b; Göbel et al., 2014). Usually, it is argued that this is due to a coactivation of verbal numerical representations such as number names when children perform symbolic numerical tasks. This activation may occur automatically, but especially children are regularly found to verbalize numerical tasks to assist processing. When no or only impaired such coactivation of verbal numerical representations and thus verbalizing is possible because of impaired language abilities (as in children with CI) this may lead to the prolonged processing times.

However, closer inspection of the performance pattern of children with CI indicated that this might only be part of the story. On the one hand, 15.6% of children with CI were found to perform significantly above average on the HRT. Interestingly, this may indicate that the normal distribution describing arithmetic performance capabilities of children with CI may be flatter and broader as compared to that of normally developing children. However, the two distributions did not differ regarding their mean as indicated by our analyses. On the other hand, poorer performance in the HRT and the observed general slowing may be associated. The HRT as used in this study is a speeded test and the observed general slowing might have led to the increased number of dyscalculia indications in the sample of children with CI. Therefore, it is inevitably necessary not only to look at these general performance impairments but also to evaluate more specific performance differences associated with particular numerical competencies and/or representations.

### SPECIFIC IMPAIRMENTS OF CHILDREN WITH CI

Evaluating the specific impairments will consider the results of the computerized calculation tasks as well as the number line estimation task. While multiplication performance has repeatedly been suggested to be closely related to language skills (Lee and Kang, 2002; Dehaene et al., 2003; see also Fazio, 1999; Koponen et al., 2006), language is generally assumed to play either no or only a minor role for solving subtractions (of either one- or multi-digit operands; McNeil and Burgess, 2002) or number line estimation (e.g., Siegler et al., 2010; but see LeFevre et al., 2010). However, even though Lee and Kang (2002) and Moeller et al. (2011) report experimental evidence for impairments of multiplication fact retrieval due to a verbal secondary task, no other specific language related impairments for multiplication have been reported in the literature. Thus, despite a generally poorer multiplication performance we did not expect any further specification of performance differences. In line with this, we observed no differences between children with CI and NH controls with regard to the processing of correctly or incorrectly solved multiplication problems as well as for the differentiation between table and non-table errors. However, closer inspection of the association of hearing age with error types revealed an interesting result with respect to non-table errors. A correlation analysis indicated that the percentage of non-table errors was correlated significantly with hearing age in children with CI [ $r(45) = -0.30, p < 0.05$ ] indicating that with increasing hearing age fewer non-table errors were committed. Interestingly, this correlation was not reliable in children without CI [ $r(49) = 0.11, n.s.$ ]. Fisher's *Z*-test indicated that the difference between the

two correlations was significant ( $Z = 1.97, p < 0.05$ ). Importantly, this pattern of results is in line with the assumption that impaired hearing/language experience of children with CI may have influenced their numerical development. Generally, multiplication is assumed to be solved via verbally mediated retrieval of arithmetic facts from long-term memory (e.g., Dehaene et al., 2003). Butterworth et al. (2003) found that with increasing skill level (i.e., automaticity of fact retrieval) the number of non-table errors decreased (see also Campbell and Graham, 1985). This is exactly what we observed for hearing/language experience of children with CI. Taken together, this indicated that the children with CI seem to process multiplication problems in a qualitatively similar way to hearing controls but with a quantitative difference arguing for an impairing influence of their reduced hearing/language experience. This interpretation was substantiated by the results for subtraction and number line estimation.

Based on the considerations of LeFevre et al. (2010; see also Helmreich et al., 2011) suggesting a link between language skills and the processing of place-value information we hypothesized that children with CI should experience particular difficulties as the demands on place-value processing increase. This is the case for (i) borrow as compared to non-borrow problems in subtraction (requiring to borrow from the tens place depending on the relation of the units) and (ii) for increasing number ranges in the number line estimation task (requiring to integrate three instead of two digits in the 0–1000 compared to the 0–100 range). As regards subtraction we observed that the borrowing effect was indeed more pronounced in children with CI as compared to NH controls. Importantly, this finding is driven by a specific decrease of performance of children with CI for subtraction problems requiring a borrow procedure. Importantly, the hypothesis that this may be associated with the impaired hearing/language experience of children with CI is corroborated when specifically considering the correlation of hearing age of these children with their performance in borrow subtraction problems. The correlation analysis revealed the to-be-expected reliable negative correlation in children with CI [ $r(45) = -0.28, p < 0.05$ , tested one-sided] indicating that with increasing hearing age borrow subtraction problems were solved faster. Moreover, this correlation was not reliable in children without CI [ $r(49) = 0.14, n.s.$ ]. And Fisher's *Z*-test indicated that the difference between the two correlations was significant ( $Z = 2.01, p < 0.05$ ). This is well in line with our argument that hearing and thus language experience of children with CI is specifically related to their place-value understanding as particularly relevant in borrow subtraction problems.

Furthermore, we also found a similar place-value related effect for the number line estimation task. With the increase of the number range from 0 to 100 to 0 to 1000 the estimation error increased more strongly for children with CI than for NH controls. Again, this supported our hypothesis of a specific impairment of children with CI that might result from poorer place-value understanding that, in turn, might originate from their atypical language development. Helmreich et al. (2011) observed that number line estimations of German-speaking children were less accurate than that of Italian-speaking children and attributed this to the way

the place-value structure of Arabic numbers is reflected in the respective languages' number words. While the order of tens and units in symbolic numbers is reflected correctly in Italian number words (e.g., 27 → ventisette, i.e., twentyseven) it is inverted in German number words (e.g., 27 → siebenundzwanzig, i.e., literally sevenandtwenty). This indicates an influence of language representations on number line estimations. In line with this LeFevre et al. (2010) found number line performance to be predicted by the linguistic pathway of their model of numerical development. In turn, this language dependency of place-value processing (e.g., Pixner et al., 2011a,b) might account for the impaired processing in children with CI when demands on place-value understanding increase in both number line estimation in a wider range and borrow subtraction problems.

Nevertheless, it is important to note that subtractions used in the present study were comprised of two-digit operands. According to the literature, language deficiencies (as seen in children with SLI) are often accompanied by working memory impairments (e.g., Henry et al., 2012). Thus, it would be plausible to speculate that children with aberrant language development such as our children with CI experience specific difficulties upon solving borrow subtraction problems because these pose heavy demands on working memory resources. Following this rationale, children with CI (and atypical language development) should be at a clear disadvantage upon solving these kinds of tasks. However, in the present study children with CI were found to have comparable verbal working memory scores and even significantly better CE scores than NH controls. Therefore, the poor working memory hypothesis cannot account for the current results. Additionally, influences of this variable have been accounted for in the analysis. Furthermore, it should be noted that in Austria, children acquire numbers up to 1000 in third grade. Because participating children attended third to fifth grade, one may assume that our results (children with CI performing poorer on large number ranges than controls) are attributable to third graders who do not yet master numbers beyond 100. However, this was not the case. Additional analyses revealed that estimation accuracy did not differ significantly between children with CI attending third grade and those attending fourth and fifth grade [ $t(43) < 0.89$ ; *n.s.*].

## CONCLUSION

Taken together, our results disclose that children with CI (and an associated atypical language development, cf. Boothroyd et al., 1991; Dawson et al., 1995; Geers et al., 2009; Pisoni et al., 2010; Ingvalson and Wong, 2013; Rinaldi et al., 2013) show general as well as specific numerical/arithmetic impairments. On the one hand, an increased number of children with CI was found to show indication of dyscalculia symptoms on a standardized arithmetical test. Additionally, children with CI were generally slower than NH controls in multiplication and subtraction. Finally, also their number line estimations were less accurate. Synced with no differences for the processing of problem and error types in multiplication this seems to indicate that the main impairment of children with CI is a general slowing. However, on the other hand, children with CI exhibited very circumscribed difficulties associated with the processing of place-value information. Performance declined more strongly than for NH controls when (i) subtraction required

a borrow procedure and (ii) number line estimation was to be performed within a wider number range requiring the integration of units, tens and hundreds instead of only units and tens. As demands on place-value understanding is increased in both of these cases the language dependency of place-value processing (LeFevre et al., 2010) might account for the impaired performance in children with CI.

It is important to note that we did not evaluate the influence of an atypical language development in children with CI on their numerical/arithmetic abilities by correlating their performance in a language task with those in numerical/arithmetic tasks. Because recent studies consistently found atypical language development in children with CI (e.g., Boothroyd et al., 1991; Dawson et al., 1995; Geers et al., 2009; Pisoni et al., 2010; Ingvalson and Wong, 2013; Rinaldi et al., 2013) we compared the performance of children with CI and a control group in specific numerical/arithmetic chosen for either their strong (i.e., multiplication fact retrieval) or weaker (i.e., two-digit subtraction and number line estimation) dependence on language-related processing. The specificity of the present results corroborated this research strategy. Yet, future studies employing a combined approach are desirable to cross-validate the present findings.

Nevertheless, our findings are original because we were able to show that a large and carefully selected group of children with CI (attending third to fifth grade) displayed overall comparable performance levels and profiles on arithmetic tasks thought to rely heavily on language demands (i.e., multiplication facts). Thus, it seems that despite initially atypical language development, which might account for the general slowing, children with CI are able to acquire arithmetic skills in a qualitatively similar fashion as their NH peers. Nonetheless, with increasing task complexity that is reflected by the necessity to quickly access the positional base-10 place-value system of the Arabic notation, children with CI perform poorer than their NH peers. The present findings have important implications for educational practice and continuing education of children with CI. In particular, our findings suggest that (i) children with CI may perform equally well than their hearing peers provided they are given some extra time to solve arithmetic problems, and (ii) children with CI may require more focused teaching of the Arabic base-10 place-value system in order to make up for their initial (and probably language-related) difficulties to acquire numerical skills.

## REFERENCES

- Adi-Bensaid, L., and Tubul-Lacy, G. (2009). Consonant free words: evidence from hebrew speaking children with cochlear implants. *Clin. Linguist. Phon.* 23, 122–132. doi: 10.1080/02699200802564961
- American Psychiatric Association [APA]. (1994). *Diagnostic, and Statistical Manual of Mental Disorders*, 4th Edn. Washington DC: American Psychiatric Association.
- Ansell, E., and Pagliaro, C. M. (2006). The relative difficulty of signed arithmetic story problems for primary level deaf and hard-of-hearing students. *J. Deaf Stud. Deaf Educ.* 11, 153–170. doi: 10.1093/deafed/enj030
- Arsalidou, M., and Taylor, M. J. (2011). Is 2+2=4? Meta-analyses of brain areas needed for numbers and calculations. *Neuroimage* 54, 2382–2393. doi: 10.1016/j.neuroimage.2010.10.009
- Badian, N. A. (1993). "Developmental dyscalculia," in *Progress in Learning Disabilities*, Vol. 5, ed. H. R. Myklebust (New York: Grune & Stratton), 235–264.

- Blarney, P. J., Barry, J. G., and Jacq, P. (2001). Phonetic inventory development in young cochlear implant users 6 years postoperation. *J. Speech Lang. Hear. Res.* 44, 73–79. doi: 10.1044/1092-4388(2001/007)
- Boothroyd, A., Geers, A. E., and Moog, J. S. (1991). Practical implications of cochlear implants in children. *Ear Hear.* 12, 81–89. doi: 10.1097/00003446-199108001-00010
- Butterworth, B., Marchesi, N., and Girelli, L. (2003). “Basic multiplication combinations: passive storage or dynamic reorganization?,” in *The Development of Arithmetic Concepts and Skills: Constructive Adaptive Expertise*, eds A. J. Baroody and A. Dowker (Mahwah, NJ: Lawrence Erlbaum Associates), 187–201.
- Campbell, J. I., and Graham, D. J. (1985). Mental multiplication skill: structure, process, and acquisition. *Can. J. Psychol.* 39, 338–366. doi: 10.1037/h0080065
- Carey, S. (2004). Bootstrapping and the origin of concepts. *Deadalus* 133, 59–68. doi: 10.1162/001152604772746701
- Cattell, R. B., Weiß, R. H., and Osterland, J. (1997). *CFT 1. Grundintelligenztest Skala 1*. Göttingen: Hogrefe.
- Chin, S. B. (2006). Realization of complex onsets by pediatric users of cochlear implants. *Clin. Linguist. Phon.* 20, 501–508. doi: 10.1080/02699200500266315
- Connor, C., Craig, H. K., Raudenbusch, S. W., Heavner, K., and Zwolan, T. A. (2006). The age at which young deaf children receive cochlear implants and their vocabulary and speech-production growth: is there an added value for early implantation? *Ear Hear.* 27, 628–644. doi: 10.1097/01.aud.0000240640.59205.42
- Dawson, P. W., Blamey, P. J., Dettman, S. J., Barker, E. J., and Clark, G. M. (1995). A clinical report in receptive vocabulary skills in cochlear implant users. *Ear Hear.* 16, 287–294. doi: 10.1097/00003446-199506000-00005
- Dehaene, S., Piazza, M., Pinel, P., and Cohen, L. (2003). Three parietal circuits for number processing. *Cogn. Neuropsychol.* 20, 487–506. doi: 10.1080/02643290244000239
- De Smedt, B., and Boets, B. (2010). Phonological processing and arithmetic fact retrieval: evidence from developmental dyslexia. *Neuropsychologia* 48, 3973–3981. doi: 10.1016/j.neuropsychologia.2010.10.018
- Dilling, H., and Freyberger, H. J. (2001). *Taschenführer zur ICD-10 Klassifikation Psychischer Störungen*. Bern: Hans Huber.
- Donlan, C., Cowan, R., Newton, E. J., and Lloyd, D. (2007). The role of language in mathematical development: evidence from children with specific language impairments. *Cognition* 103, 23–33. doi: 10.1016/j.cognition.2006.02.007
- Fazio, B. B. (1999). Arithmetic calculation, short-term memory, and language performance in children with specific language impairment: A 5-yr follow up. *J. Speech Lang. Hear. Res.* 42, 420–431. doi: 10.1044/jslhr.4202.420
- Feigenson, L., Dehaene, S., and Spelke, E. (2004). Core systems of number. *Trends Cogn. Sci.* 8, 307–314. doi: 10.1016/j.tics.2004.05.002
- Friedmann, N., and Sztzman, R. (2011). The comprehension and production of Wh-Questions in deaf and hard-of-hearing children. *J. Deaf Stud. Deaf Educ.* 16, 212–235. doi: 10.1093/deafed/enq052
- Geers, A. E., Moog, J. S., Biedenstein, J., Brenner, C., and Hayes, H. (2009). Spoken language scores of children using cochlear implants compared to hearing age-mates at school entry. *J. Deaf Stud. Deaf Educ.* 14, 371–385. doi: 10.1093/deafed/enn046
- Geers, A. E., Nicholas, J. G., and Sedey, A. L. (2003). Language skills of children with early cochlear implantation. *Ear Hear.* 24, 46–58. doi: 10.1097/01.AUD.0000051689.57380.1B
- Göbel, S., Moeller, K., Pixner, S., Kaufmann, L., and Nuerk, H. C. (2014). Language affects symbolic arithmetic in children: the case of number word inversion. *J. Exp. Child Psychol.* 119, 17–25. doi: 10.1016/j.jecp.2013.10.001
- Göbel, S., and Snowling, M. J. (2010). Number processing skills in adults with dyslexia. *Q. J. Exp. Psychol.* 63, 1361–1373. doi: 10.1080/17470210903359206
- Gross-Tsur, V., Manor, O., and Shalev, R. S. (1996). Developmental dyscalculia: prevalence and demographic features. *Dev. Med. Child Neurol.* 38, 25–33. doi: 10.1111/j.1469-8749.1996.tb15029.x
- Haffner, J., Baro, K., Parzer, P., and Resch, F. (2005). *Heidelberger Rechentest 1-4*. Göttingen: Hogrefe.
- Helmreich, I., Zuber, J., Pixner, S., Kaufmann, L., Nuerk, H.-C., and Moeller, K. (2011). Language effects on children’s nonverbal number line estimations. *J. Cross Cult. Stud.* 42, 598–613. doi: 10.1177/0022022111406026
- Henry, L. A., Messer, D. J., and Nash, G. (2012). Executive functioning in children with specific(language)impairment. *J. Child Psychol. Psychiatry* 53, 37–45. doi: 10.1111/j.1469-7610.2011.02430.x
- Ingvalson, E. M., and Wong, P. C. M. (2013). Training to improve language outcomes in cochlear implant recipients. *Front. Psychol.* 4:263. doi: 10.3389/fpsyg.2013.00263
- Johnson, C., and Goswami, U. (2010). Phonological awareness, vocabulary, and reading in deaf children with cochlear implants. *J. Speech Lang. Hear. Res.* 53, 237–261. doi: 10.1044/1092-4388(2009/08-0139)
- Klein, E., Moeller, K., Dressel, K., Domahs, F., Wood, G., Willmes, K., et al. (2010a). To carry or not to carry – is this the question? Disentangling the carry effect in multi-digit addition. *Acta Psychol.* 135, 67–67. doi: 10.1016/j.actpsy.2010.06.002
- Klein, E., Willmes, K., Dressel, K., Domahs, F., Wood, G., Nuerk, H.-C., et al. (2010b). Categorical and continuous – disentangling the neural correlates of the carry effect in multi-digit addition. *Behav. Brain Funct.* 6, 70. doi: 10.1186/1744-9081-6-70
- Koponen, T., Mononen, R., Räsänen, P., and Ahonen, T. (2006). Basic numeracy in children with specific language impairment: heterogeneity and connections to language. *J. Speech Lang. Hear. Res.* 49, 58–73. doi: 10.1044/1092-4388(2006/005)
- Kramer, F. (2007). *Kulturfaire Berufseignungsdiagnostik bei Gehörlosen und Daraus Abgeleitete Untersuchungen zu den Unterschieden der Rechenfertigkeiten bei Gehörlosen und Hörenden*. Ph.D. thesis, Universität Aachen, Aachen.
- Lee, K. M., and Kang, S. Y. (2002). Arithmetic operation and working memory: differential suppression in dual tasks. *Cognition* 83, B63–B68. doi: 10.1016/S0010-0277(02)00010-0
- LeFevre, J. A., Fast, L., Skwarchuk, S.-L., Smith-Chant, B. L., Bisanz, J., Kamawar, D., et al. (2010). Pathways to mathematics: longitudinal predictors of performance. *Child Dev.* 81, 1753–1767. doi: 10.1111/j.1467-8624.2010.01508.x
- Leyrer, M. (2008). *Die prosodische Entwicklung nach Cochlea-Implantation: Eine Fallstudie*. Saarbrücken: VDM Verlag.
- McNeil, J. E., and Burgess, P. W. (2002). The selective impairment of arithmetical procedures. *Cortex* 38, 569–587. doi: 10.1016/S0010-9452(08)70023-6
- Miles, T. R., Haslum, M. N., and Wheeler, T. J. (2001). The mathematical abilities of dyslexic 10-years-olds. *Ann. Dyslexia* 51, 299–321. doi: 10.1007/s11881-001-0015-0
- Moeller, K., Klein, E., Fischer, M. H., Nuerk, H.-C., and Willmes, K. (2011). Representation of multiplication facts – evidence for partial verbal coding. *Behav. Brain Funct.* 7, 25. doi: 10.1186/1744-9081-7-25
- Moeller, K., Pixner, S., Kaufmann, L., and Nuerk, H.-C. (2009). Children’s early mental number line: logarithmic or rather decomposed linear? *J. Exp. Child Psychol.* 103, 503–515. doi: 10.1016/j.jecp.2009.02.006
- National Council of Teachers of the Deaf Research Committee. (1957). The teaching of arithmetic in schools for the deaf. *Teach. Deaf* 151, 72–165.
- Nott, P., Cowan, R., Brown, P. M., and Wigglesworth, G. (2003). Assessment of language skills in young children with profound hearing loss under two years of age. *J. Deaf Stud. Deaf Educ.* 8, 401–421. doi: 10.1093/deafed/eng030
- O’Hearn, K., and Luna, B. (2009). Mathematical skills in Williams syndrome: insight into the importance of underlying representations. *Dev. Disabil. Res. Rev.* 15, 11–20. doi: 10.1002/ddrr.47
- Pisoni, D. B., Cinway, C. M., Kronenberger, W., Henning, S., and Anaya, E. (2010). “Executive function, cognitive control, and sequence learning in deaf children,” in *The Oxford Handbook of Deaf Studies, Language, and Education with Cochlear Implants*, eds M. Marshark and P.E. Spencer (New York: Oxford University Press), 439–457.
- Pixner, S., Moeller, K., Hermanova, V., Nuerk, H.-C., and Kaufmann, L. (2011a). Whorf-reloaded: language effects on nonverbal number processing in first grade. A trilingual study. *J. Exp. Child Psychol.* 108, 371–382. doi: 10.1016/j.jecp.2010.09.002
- Pixner, S., Zuber, J., Hermanova, V., Kaufmann, L., Nuerk, H.-C., and Moeller, K. (2011b). One language, two number-word system and many problems: numerical cognition in the czech language. *Res. Dev. Disabil.* 32, 2683–2689. doi: 10.1016/j.ridd.2011.06.004
- Rinaldi, P., Baruffandi, F., Burdo, S., and Caselli, M. C. (2013). Linguistic and pragmatic skills in toddlers with cochlear implant. *Int. J. Lang. Commun. Disord.* 48, 715–725. doi: 10.1111/1460-6984.12046
- Siegler, R. S., and Booth, J. L. (2004). Development of numerical estimation in young children. *Child Dev.* 75, 428–444. doi: 10.1111/j.1467-8624.2004.00684.x

- Siegler, R. S., Thompson, C. A., and Schneider, M. (2010). An integrated theory of whole number and fraction development. *Cogn. Psychol.* 62, 273–296. doi: 10.1016/j.cogpsych.2011.03.001
- Simmons, F. R., and Singleton, C. (2008). Do weak phonological representations impact on arithmetic development? A review of research into arithmetic and dyslexia. *Dyslexia* 14, 77–94. doi: 10.1002/dys.341
- Snowling, M. J. (2000). *Dyslexia*. Malden, MA: Blackwell Publishers.
- Tomblin, J. B., Barker, B. A., Spencer, L. J., Zhang, X., and Gantz, B. J. (2005). The effect of age at cochlear implant initial stimulation on expressive language growth in infants and toddlers. *J. Speech Lang. Hear. Res.* 48, 853–867. doi: 10.1044/1092-4388(2005/059)
- von Aster, M. G., Schweitzer, M., and Weinhold Zulauf, M. (2007). Rechenstörungen bei Kindern: vorläufer, prävalenz und psychische symptome. *Z. Entwicklungspsychol. Pädagog. Psychol.* 39, 85–96. doi: 10.1026/0049-8637.39.2.85
- von Aster, M. G., and Shalev, R. (2007). Number development and developmental dyscalculia. *Dev. Med. Child Neurol.* 49, 868–873. doi: 10.1111/j.1469-8749.2007.00868.x
- Weiß, R. H. (2008). *Grundintelligenztest Skala 2 Revision (CFT 20 R)*. Göttingen: Hogrefe.
- Zarfaty, Y., Nunes, T., and Bryant, P. (2004). The performance of young deaf children in spatial and temporal number tasks. *J. Deaf Stud. Deaf Educ.* 9, 315–326. doi: 10.1093/deafed/enh034
- Zuber, J., Pixner, S., Moeller, K., and 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. *J. Exp. Child Psychol.* 102, 60–77. doi: 10.1016/j.jecp.2008.04.003

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 08 July 2014; accepted: 01 December 2014; published online: 16 December 2014.

Citation: Pixner S, Leyrer M and Moeller K (2014) Number processing and arithmetic skills in children with cochlear implants. *Front. Psychol.* 5:1479. doi: 10.3389/fpsyg.2014.01479

This article was submitted to *Developmental Psychology*, a section of the journal *Frontiers in Psychology*.

Copyright © 2014 Pixner, Leyrer and Moeller. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Contribution of working memory in multiplication fact network in children may shift from verbal to visuo-spatial: a longitudinal investigation

Mojtaba Soltanlou<sup>1,2,3</sup>, Silvia Pixner<sup>4</sup> and Hans-Christoph Nuerk<sup>2,3,5\*</sup>

<sup>1</sup> Graduate Training Centre of Neuroscience/International Max Planck Research School for Cognitive and Systems Neuroscience, Tuebingen, Germany, <sup>2</sup> Knowledge Media Research Center, Tuebingen, Germany, <sup>3</sup> Department of Psychology, Eberhard Karls University, Tuebingen, Germany, <sup>4</sup> Institute of Applied Psychology, UMIT – The Health and Life Sciences University, Hall in Tyrol, Austria, <sup>5</sup> LEAD Graduate School, Eberhard Karls University, Tuebingen, Germany

## OPEN ACCESS

### Edited by:

Yvette Renee Harris,  
Miami University, USA

### Reviewed by:

Anja Ischebeck,  
University of Graz, Austria  
Alex M. Moore,  
University of Missouri, USA

### \*Correspondence:

Hans-Christoph Nuerk,  
Department of Psychology, Eberhard  
Karls University, Schleichstrasse 4,  
Tuebingen 72076, Germany  
hc.nuerk@uni-tuebingen.de

### Specialty section:

This article was submitted to  
*Developmental Psychology*,  
a section of the journal  
*Frontiers in Psychology*

**Received:** 06 December 2014

**Accepted:** 10 July 2015

**Published:** 23 July 2015

### Citation:

Soltanlou M, Pixner S and Nuerk H-C  
(2015) Contribution of working  
memory in multiplication fact network  
in children may shift from verbal  
to visuo-spatial: a longitudinal  
investigation.  
*Front. Psychol.* 6:1062.  
doi: 10.3389/fpsyg.2015.01062

Number facts are commonly assumed to be verbally stored in an associative multiplication fact retrieval network. Prominent evidence for this assumption comes from so-called operand-related errors (e.g.,  $4 \times 6 = 28$ ). However, little is known about the development of this network in children and its relation to verbal and non-verbal memories. In a longitudinal design, we explored elementary school children from grades 3 and 4 in a multiplication verification task with the operand-related and -unrelated distractors. We examined the contribution of multiplicative fact retrieval by verbal and visuo-spatial short-term and working memory (WM). Children in grade 4 showed smaller reaction times in all conditions. However, there was no significant difference in errors between grades. Contribution of verbal and visuo-spatial WM also changed with grade. Multiplication correlated with verbal WM and performance in grade 3 but with visuo-spatial WM and performance in grade 4. We suggest that the relation to verbal WM in grade 3 indicates primary linguistic learning of and access to multiplication in grade 3 which is probably based on verbal repetition of the multiplication table heavily practiced in grades 2 and 3. However, the relation to visuo-spatial semantic WM in grade 4 suggests that there is a shift from verbal to visual and semantic learning in grade 4. This shifting may be induced because later in elementary school, multiplication problems are rather carried out via more written, i.e., visual tasks, which also involve executive functions. More generally, the current data indicates that mathematical development is not generally characterized by a steady progress in performance; rather verbal and non-verbal memory contributions of performance shift over time, probably due to different learning contents.

**Keywords:** multiplication, arithmetic, fact retrieval, operand errors, verbal working memory, visuo-spatial working memory

## Introduction

Children usually get better in arithmetic problem solving with age and experience. For instance, the processing strategy of multiplication in children changes from procedure- and strategy-based calculation to retrieval during developmental ages (Cooney et al., 1988; Lemaire and Siegler, 1995). It has been reported that there is a transition to retrieval process for solving single-digit multiplication problems in grade 4 (Cooney et al., 1988). However, this retrieval process is not constant during the following years of development (Campbell and Graham, 1985). Nonetheless, longitudinal studies for verification of this claim are scarce. In particular, the development of the automatic associations within the fact retrieval network has not been sufficiently understood.

Of major importance in multiplication verification performance is operand-relatedness. Operand-relatedness is whether the presented or responded answer belongs to the table of one of the operands or not. For instance, in a production task, an operand-related error is when a participant responds with 24 when presented with the problem  $7 \times 4$  because 24 is part of the same multiplication table of one of the operands (here the 4). An operand-unrelated error would be the solution 30 because this number belongs neither to the multiplication table of 4 nor of 7. In a verification task for the problem  $4 \times 6 = 24$ , an operand-related verification distractor would be  $4 \times 6 = 28$ , and the operand-unrelated distractor would be  $4 \times 6 = 29$ .

It has been reported that the operand-related distractor errors make up about 87.5% of all errors in adults (Campbell, 1997; Domahs et al., 2006) and about 75.7% of all errors in children (Butterworth et al., 2003). The large frequency of operand-related errors has been explained in terms of a developing memory representation in an interrelated network of facts (Ashcraft, 1987). This representation means that during retrieval of a multiplication answer from an interconnected multiplication network, the operand-related distractors will activate the retrieval processing more than the operand-unrelated distractors and lead to a slower response with more errors. These assumptions have been implemented in the network interference model which explains that arithmetic facts are stored as nodes in an associative network in long-term memory and are retrieved via a spreading activation (Campbell, 1995). The presented multiplication generates activation in the corresponding nodes and this activation spreads along the connecting pathways to associated nodes. For example, the presentation of  $7 \times 3$  activates node 7 along with its related nodes (14, 21, 28, etc.) and node 3 with its related nodes (6, 9, 12, etc.). In other terms, the activation of associates which are the operand-related distractors (e.g., 28 instead of 21 in the example above), increases the accessibility of these associates. Consequently, it is more plausible to verify it erroneously as a correct answer. However, in the operand-unrelated distractors (e.g., 25 instead of 21 in the example above), there is minimum activation of the associates, hereby decreasing the accessibility of them as a correct answer. Hence, activation of multiple associates interferes with the solutions because it renders these associates more accessible.

To our knowledge, there are very few longitudinal studies in regard to multiplication development in children considering operand-relatedness. For instance, in a study by Lemaire and Siegler (1995) it was shown that in three sessions of multiplication production assessment in grade 2, the proportion of both operand-related and -unrelated errors increased. The other study which used multiplication verification in children, did not report error analyses because it was stable at about 6% in grades 3 and 4 (De Brauwer and Fias, 2009). Therefore, it is still unclear if error patterns and their relation to operand-relatedness change longitudinally in children and consequently what can be inferred with regard to the longitudinal change in the multiplication fact retrieval network.

From the structure of the network interference model, two hypotheses could be brought forward for our longitudinal developmental study on multiplication facts. (i) Because the strength of the association network could increase with age and experience, the operand-relatedness error effect should be larger in older children. (ii) The alternative hypothesis would be that the network becomes more refined in reciprocal inhibition so that the single entries can be better separated with age and experience. Then, the operand-relatedness error effect should be smaller in older children. In our opinion, both views are possible. The current study set out to discern these two hypotheses.

Another main issue of this study is that to our knowledge the possible varying influence of other cognitive processes on the multiplication performance has not been studied longitudinally in children. One natural candidate for such a cognitive process is memory, containing working memory (WM) and short-term memory (STM). One account of WM capacity is defined by Shah and Miyake (1996) and Miyake and Shah (1999). In this model WM capacity contains two separate pools of domain-specific resources for verbal and visuo-spatial information. Each domain keeps and manipulates information independently from the other. This distinction between verbal and visuo-spatial domains has been supported by the previous findings (e.g., Friedman and Miyake, 2000; Miyake et al., 2001; Jarvis and Gathercole, 2003). WM has been reported as a pure measure of a child's learning potential (Alloway and Alloway, 2010). Thus, it has been assumed to predict a child's performance in mathematic learning based on the WM skills (Alloway and Passolunghi, 2011). While WM is defined as an ability of storage and manipulation of information, STM is considered as only storage of information for a temporary period of time (for more see Alloway et al., 2006). In other words, WM is a memory system containing separable interacting components, while STM is almost a single store (Alloway et al., 2006). In sum, STM demonstrates temporal deterioration and capacity limits, whereas WM is a multi-component system that stores and manipulates information in STM and uses attention to manage STM and applies STM to cognitive tasks (Baddeley and Hitch, 1974; Cowan, 1988; Baddeley, 1992, 1998; for more see Cowan, 2008). Therefore, STM involves a minimal load of processing, while WM contains an additional process for manipulation of information that leads to higher loading of process. Different components of STM and WM have already been reported to

be involved in different mathematical tests during developing stages (see also Meyer et al., 2010) but the possibility of their different role in development of multiplication has not been longitudinally considered – therefore, the differential roles of STM and WM will also be considered in the current study.

Recent studies have shown that the relative contributions of memory components to general mathematic learning changes during development ages. At first, preschool children rely more on visuo-spatial memory than verbal memory for learning and remembering arithmetic; therefore, the best predictor of the arithmetic performance at this age is visuo-spatial sketchpad capacity (McKenzie et al., 2003; Simmons et al., 2008). Later, starting from school age, learning is more dependent on verbal rehearsal to preserve information in memory, thus recruiting more the phonological loop (Hitch et al., 1988; Rasmussen and Bisanz, 2005). This has been explained by verbally mediated strategies, in which children transform symbols and numbers into verbal code (Logie et al., 1994; Geary et al., 1996). By the first grade, performance relies equivalently on non-verbal and verbal memory. Meyer et al. (2010) showed that the verbal components of memory predict mathematical reasoning skill in grade 2, whereas the visuo-spatial component is the predictor in grade 3. Therefore, different WM and STM components seem to be critical for mathematics learning in general. However, currently we have only little data on how the different verbal and visuo-spatial components of WM and STM contribute to multiplication performance in different ages in elementary school and how the importance of such components changes over time. For our study, we hypothesized a shift between memory components, from verbal to visuo-spatial, in children during development in multiplication similarly to those reported by Meyer et al. (2010) for mathematical reasoning. In the current study as we collected longitudinal data, the first aim was to evaluate in which way children process multiplication in grades 3 and 4. According to the previous findings, we expected children in grade 4 to be faster and possibly less error-prone than in grade 3. The second aim was to investigate whether their memory processing is differentially influenced by operand-relatedness with age and experience, especially with regard to the error data. Finally, the third and main aim of this study was to investigate the contributions of verbal linguistic and visuo-spatial non-verbal representations on arithmetic skill, namely the influence of verbal and visuo-spatial STM and WM on multiplication skill.

## Materials and Methods

The current study was part of a large longitudinal project evaluating numerical development from grade 1 to grade 4. In this study, we focused on the development of multiplication performance which was measured only from grade 3 to grade 4.

## Participants

In total, 77 native German-speaking Austrian children (39 girls and 38 boys) were assessed in multiplication both at the end of

grades 3 and 4. The children were between 8 years 6 months and 10 years 5 months ( $M = 9$  years 4 months,  $SD = 7$  months) in grade 3 and 1 year older in grade 4. All children had normal or corrected-to-normal vision and IQ scores in the normal range. No child received special education services or had documented brain injury or behavioral problems. This study was carried out in accordance with the recommendations of the Landesschulrat, the regional school administration, which was responsible for approval of school-related studies in Austria at that time. Parents of all subjects gave written informed consent in accordance with the Declaration of Helsinki.

## Multiplication Stimuli

Children were tested on a computerized multiplication verification task. The experiment started with eight practice trials. Multiplication problems (range of operands: 3–8; problem size: 13–54) along with the answer probe were presented at the same time on the screen in white against a black background (font: Arial; size: 48-point). Problems were presented in the form  $x \times x = xx$  at the  $x/y$  coordinates (512/300) on a screen with the resolution set to  $1024 \times 768$ . In total there were 80 multiplication trials. Half of the trials were true (i.e., the solutions were displayed) and half of them were false (i.e., distractors which had to be rejected were displayed). The distractors consisted of operand-related and operand-unrelated trials. In the operand-related trials the operand split was  $\pm 1$  from the solutions on the multiplication table (e.g.,  $6 \times 3 = 21$ ). In the operand-unrelated trials the displayed answers were not from the multiplication table. In the operand-unrelated trials the displayed answer differed from the solution by  $\pm 2$  to  $\pm 9$ , with the average split matched at 0.4 (e.g.,  $6 \times 3 = 13$ ). The task was a verification paradigm where the displayed answer needed to be verified as correct or incorrect. Problem size was held approximately constant between item categories. Problems and answer probes were presented until a response was given or the response time (RT) of 15000 ms finished. The response was made by pressing the “Alt” or “Alt Gr” button of a QWERTZ keyboard to verify whether the displayed answer was the solution or distractor, respectively. It is essential to note that the solutions and distractors refer to the stimuli presented in the verification task, not the children’s responses. The children’s responses were correct or incorrect. The fixation cross was presented at the beginning of each trial for 500 ms. The inter-stimulus interval was set to 1500 ms. No feedback was given.

## Memory Tasks

Four memory components including verbal and visuo-spatial STM and verbal and visuo-spatial WM (Alloway et al., 2006; Alloway and Passolunghi, 2011) were assessed in the present study. For verbal STM, children were asked to immediately recall spoken sequences of letters (presentation rate: one letter per second). Starting with two-item sequences, sequence length was increased by one letter when at least two of three given sequences were recalled correctly; otherwise, testing was stopped. The verbal STM score was the maximum sequence length at which at least two sequences were repeated correctly. For visuo-spatial STM,

in a block tapping task (Corsi, 1973), children needed to repeat pointing to cubes in the same order as the experimenter. Again, children started with two-item sequences. The procedure and scoring were identical to those in letter repetition. In general, forward span tests were defined as STM and backward span tests were defined as WM (Cowan, 1988; see also Cowan, 2008).

For verbal and visuo-spatial WM, children were asked to recall sequences of letters and blocks in reverse order. The procedure and scoring were identical to those in the STM tasks. It is noteworthy that the current study included forward recall as a measure of verbal and visuo-spatial STM and backward recall as a measure of verbal and visuo-spatial WM. In forward recall tasks the processing load is minimal as children immediately recall the sequences (Alloway et al., 2006). In contrast, in the backward recall tasks there is an additional requirement to recall the reverse sequence that imposes a substantial processing load on the child. This higher processing load has been illustrated by the finding that forward spans scores are higher than backward spans (Isaacs and Vargha-Khadem, 1989; Vandierendonck et al., 2004).

## Procedure

All children were assessed individually in one-on-one sessions in a separate room. In both grades, multiplication performance and short-term and WM were assessed.

## Analysis

Response times were measured by key-press. Only RTs for correct responses were entered into the analyses. Furthermore, response latencies shorter than 200 ms or longer than 15000 ms were not considered; however, there was no response out of this range. In a second step, responses outside the interval of  $\pm 3$  SD around the individual mean were excluded. Thus, about 3% of the responses in grade 3 and about 4.5% of the responses in grade 4 were not considered for further analyses. First, we ran two repeated-measures analyses of variance (ANOVAs), first for the solution and distractor (operand-related and -unrelated together) trials for both grades and second for the operand-related and operand-unrelated distractors for both grades. Second, the correlation of the WM components was analyzed using stepwise multiple linear regression analysis on mean RTs and error rates. For the error analysis, an arcsine-square-root transformation was applied to approximate normal distribution (e.g., Winer et al., 1971).

Because of controversies regarding confirmation of null hypothesis using traditional statistical inference, the Bayesian method was used in the current study. The method described in detail by Masson (2011) enables calculating graded evidence for null hypothesis (i.e., no difference between groups) and alternative hypothesis (i.e., difference between groups). In the analysis, sum of squares and number of observations from ordinal ANOVA were used to calculate Bayesian factors which then can be used to calculate posterior probabilities (see also Raftery, 1995). In fact, we employed the Bayesian method in order to estimate the likelihood of correctness of the null and alternative hypotheses.

## Results

Trials with RTs 3 SDs above or below a child's average RT were excluded. Children with trial exclusion or an error rate of more than 33% were not considered [six children (mean age = 9 years 4 months, two girls and four boys)]. Thus, the data of 71 children was considered in the analyses. Children had on average significantly higher WM scores in grade 4 than in grade 3 (see **Table 1**). A previous study suggested that the window between second and third grades is too short a time frame for major changes in WM capacity (Meyer et al., 2010) but interestingly we found that this difference is statistically significant between grades 3 and 4.

## Solution vs. Distractor

First, we investigated the effect of grade on the solution and distractor (both operand-related and -unrelated together) trials for RTs and accuracy.

## Response Times

Raw RT of correct responses was analyzed by repeated-measures ANOVA with grade (3 or 4) and condition (solution or distractor) as within-participant factors. Children took on average 3118 ms (SD = 1243 ms) to choose the correct answer in grade 3 and 2320 ms (SD = 916 ms) in grade 4. Children in grade 4 were on average 798 ms faster than in grade 3,  $F(1,70) = 58.46$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.46$ . RTs for the solution condition was 531 ms faster than for the distractor condition which indicated a significant difference between the two conditions,  $F(1,70) = 162.07$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.70$ . Interaction of grade  $\times$  condition showed that the effect of grade is greater for the distractor than for the solution,  $F(1,70) = 9.14$ ,  $p = 0.003$ ,  $\eta_p^2 = 0.12$  (**Figure 1A**; **Table 2**). Bayesian analysis revealed that the posterior probability of null hypothesis for grade and condition was about zero (the same probability of alternative hypothesis was complementary, i.e., about 1). The posterior probability of null hypothesis for interaction was 0.10 (the same probability of alternative hypothesis was 0.90).

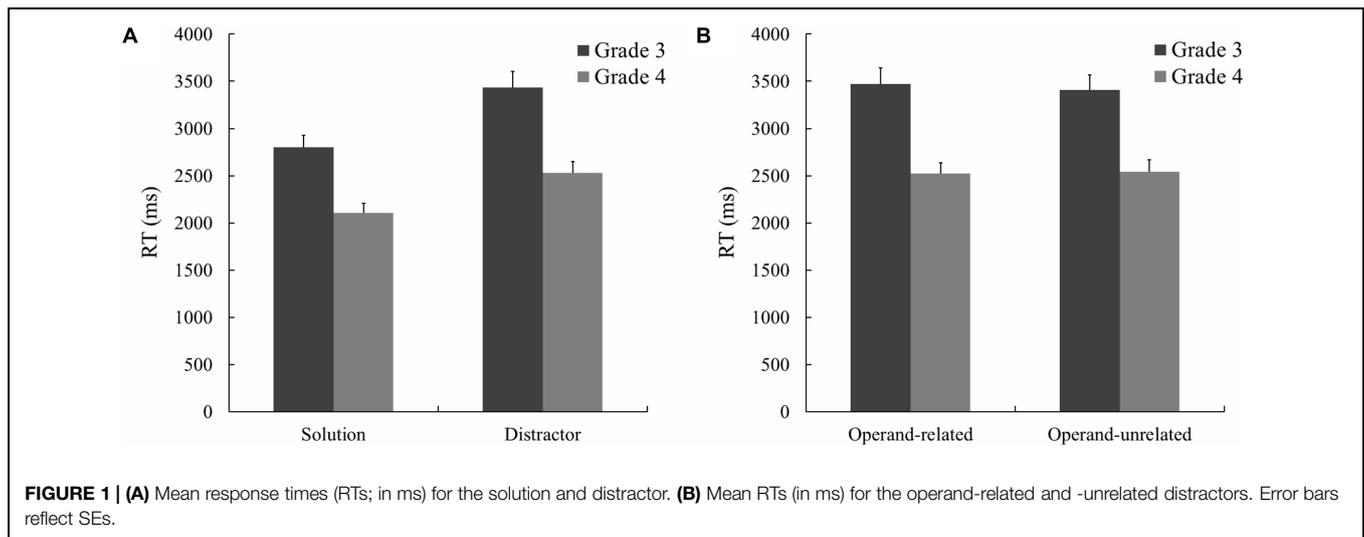
## Error Rates

Error rates were analyzed by repeated-measures ANOVAs with grade (3 or 4) and condition (solution or distractor) as within-participant factors. Overall, children responded incorrectly on 6.11% of all trials in grade 3 and on 6.51% in grade 4. Error rates did not differ significantly neither between the

**TABLE 1 | Means and SDs of memory components.**

Variable	Grade 3		Grade 4		$t^a$	$p^b$
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
Verbal short-term memory (STM)	4.55	0.73	4.92	0.73	-4.68	<0.001
Verbal working memory (WM)	2.89	0.60	3.30	0.55	-4.72	<0.001
Visuo-spatial STM	5.06	0.70	5.56	0.67	-4.88	<0.001
Visuo-spatial WM	4.18	1.10	4.69	0.86	-3.82	<0.001

<sup>a</sup>Paired sample *t* test. <sup>b</sup>Two-tailed significance level of 0.01.



**TABLE 2 | Mean response times (RTs) and error rates (and SDs) for multiplication trials.**

		Grade 3		Grade 4	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
RT (ms)	Solution	2799	1091	2108	847
	Operand-related distractor	3468	1466	2523	948
	Operand-unrelated distractor	3406	1371	2544	1045
Errors (%)	Solution	6.30	6.24	5.77	6.04
	Operand-related distractor	7.68	9.41	8.94	10.52
	Operand-unrelated distractor	4.15	7.37	5.56	8.17

grades,  $F(1,70) = 0.11$ ,  $p = 0.74$ ,  $\eta_p^2 = 0.002$ , between the conditions,  $F(1,70) = 0.095$ ,  $p = 0.76$ ,  $\eta_p^2 = 0.001$ , nor in their interaction,  $F(1,70) = 3.04$ ,  $p = 0.09$ ,  $\eta_p^2 = 0.042$ . Thus, the RT differences could not be explained by speed-accuracy trade-offs. Bayesian analysis revealed that the posterior probability of null hypothesis for grade and condition was 0.89 (the same probability of alternative hypothesis was 0.11). The posterior probability of null hypothesis for interaction was 0.65 (the same probability of alternative hypothesis was 0.35). This is rated as positive evidence for the null hypothesis applying the criteria suggested by Masson (2011).

### Operand-Related vs. Operand-Unrelated

Second, we investigated the effect of grade on the operand-related and operand-unrelated distractor trials for RTs and accuracy. Note that this analysis was done for the distractors only.

### Response Times

Raw RT of correct responses was analyzed by repeated-measures ANOVA with grade (3 or 4) and condition (operand-related or operand-unrelated) as within-participant factors. Children in grade 4 were on average 903 ms faster than in grade 3,  $F(1,70) = 53.74$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.43$ . Raw RT neither differed significantly between conditions,  $F(1,70) = 0.28$ ,  $p = 0.60$ ,

$\eta_p^2 = 0.004$ , nor did interaction between conditions and grade,  $F(1,70) = 1.57$ ,  $p = 0.22$ ,  $\eta_p^2 = 0.022$ , (Table 2; Figure 1B). Bayesian analysis revealed that the posterior probability of null hypothesis for grade was about zero (the same probability of alternative hypothesis was about 1). However, the posterior probability of null hypothesis for condition was 0.88 (the same probability of alternative hypothesis was 0.12); and for interaction 0.79 (the same probability of alternative hypothesis was 0.21).

### Error Rates

Error rates were analyzed by repeated-measures ANOVAs with grade (3 or 4) and condition (operand-related or operand-unrelated) as within-participant factors. The operand-related distractor trials were significantly more error-prone than the operand-unrelated distractor,  $F(1,70) = 22.82$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.25$ . Error rates neither differed significantly between the grades,  $F(1,70) = 1.43$ ,  $p = 0.24$ ,  $\eta_p^2 = 0.02$ , nor did interaction between conditions and grade,  $F(1,70) = 0.06$ ,  $p = 0.81$ ,  $\eta_p^2 = 0.001$ . Bayesian analysis revealed that the posterior probability of null hypothesis for grade was 0.80 (the same probability of alternative hypothesis was 0.20). However, the posterior probability of null hypothesis for condition was about zero (the same probability of alternative hypothesis was about 1); and for interaction 0.89 (the same probability of alternative hypothesis was 0.11).

### Relation between Multiplication Performance and Memory Components

#### Regression Analysis<sup>1</sup>

In order to investigate which memory component predicted multiplication performance in grades 3 and 4, a series of

<sup>1</sup>We know from many previous numerical and arithmetic experiments that RT data in children are very noisy. Hence, employing z-transformed RT to reduce inter-individual differences in intra-individual variance (cf. Nuerk et al., 2004, and many following papers since), we reanalyzed linear regressions. In general, none of the memory components predicted z-transformed RTs in grade 3. In grade 4 the verbal WM component predicted solution z-transformed RT, distractor z-transformed RT, and operand-related distractor z-transformed RT. However, this

stepwise regression analyses were conducted. For each grade, one regression predicted each of the 10 verification dependent variables (total RT, solution RT, distractor RT, operand-related distractor RT, operand-unrelated distractor RT, total error, solution error, distractor error, operand-related distractor error, and operand-unrelated distractor error) from the four memory components measured concurrently. All four memory scores were entered simultaneously with a stepwise function. This approach allowed us to identify the best predictors for different dependent variables in both grades. The model of total errors in grade 3 comprised only the predictor verbal WM,  $R^2 = 0.057$ , adjusted  $R^2 = 0.044$ ,  $F(1,69) = 4.193$ ,  $p = 0.044$ , while the other memory components failed to explain significant amounts of additional variance. Inspection of the individual beta weights indicated a significant influence of verbal WM (Table 3). The model of the operand-unrelated distractor errors in grade 3 comprised only the predictors verbal WM and verbal STM,  $R^2 = 0.178$ , adjusted  $R^2 = 0.153$ ,  $F(2,68) = 7.340$ ,  $p = 0.001$ , while the other memory components failed to explain significant amounts of additional variance. Inspection of the individual beta weights indicated a significant influence of verbal WM and verbal STM (Table 3). The model of total errors in grade 4 comprised only the predictor visuo-spatial WM,  $R^2 = 0.072$ , adjusted  $R^2 = 0.058$ ,  $F(1,69) = 5.325$ ,  $p = 0.024$ , while the other memory components failed to explain significant amounts of additional variance. Inspection of the individual beta weights indicated a significant influence of visuo-spatial WM (Table 3). All other predictors and criterion variables were not significant in regression analyses. Bayesian analysis revealed that the posterior probability of null hypothesis for total error in grade 3 was 0.51 (the same probability of alternative hypothesis was 0.49). However, the posterior probability of null hypothesis for the operand-related distractor error was about zero (the same probability of alternative hypothesis was about 1); and for total error in grade 4 0.38 (the same probability of alternative hypothesis was 0.62).

## Discussion

In the current study we collected longitudinal data from children in grades 3 and 4. The first aim of the study was to evaluate how children process multiplication in different grades. The second aim was to investigate the development of the multiplication fact retrieval network, i.e., whether their memory of multiplication

suggests that intra-individual noise in the RT data may at least partially account for the null effects observed in RTs.

facts is influenced by operand-relatedness. Furthermore, the third and main aim of this study was to investigate the contributions of verbal and visuo-spatial STM and WM to multiplication skill.

## Multiplication Fact Fluency Increases Longitudinally with Age and Experience

As we expected, children in grade 4 were faster than in grade 3 which is in line with previous findings that children become faster during development (Koshmider and Ashcraft, 1991; Lemaire et al., 1996; Butterworth et al., 2003; De Brauwer and Fias, 2009). Although children in both grades depended heavily on memory retrieval to solve the simple one-digit problems, this retrieval processing was more dominant in grade 4 (Verguts and Fias, 2005). Thus, because of the faster processing, verification of the solution, and rejection of the distractor was faster.

As regards RTs, children in both grades verified the solutions faster than the distractors (Koshmider and Ashcraft, 1991; De Brauwer and Fias, 2009). Koshmider and Ashcraft (1991) explained this result by saying that the solutions facilitate verification of the correct answer in children when the solutions are used as a prime, probably because the solutions make the strongest activation in the related nodes which in turn accelerates memory retrieval process.

As regards errors, the difference of error rate between the solutions and distractors was not statistically significant in the current study: the error rates remained stable, about 6% in grades 3 and 4. Again, this non-significant change in error rates is in line with previous results (Koshmider and Ashcraft, 1991; De Brauwer and Fias, 2009).

In brief, children in grade 4 were faster in both conditions than in grade 3 but their performance in regard to error did not differ significantly. This can be explained by more efficient and faster solving strategies with age which are, however, not yet more accurate than the slower strategies of younger children.

## No Changes in the Operand-Relatedness Effect with Age and Experience

In line with our main hypothesis, the operand-related distractors were erroneously responded to significantly more frequently than the operand-unrelated distractors. The finding is in line with the previous studies in children (Koshmider and Ashcraft, 1991; Lemaire and Siegler, 1995; Butterworth et al., 2003) which reported operand-related errors as the most frequent errors. It implies that multiplication facts are stored in the associative network already 1 year after the first multiplication facts are

**TABLE 3 | Results for significant predictors entered in the stepwise multiple regression analysis.**

Grade	Variable	Predictor	B	Standardized $\beta$	t	p
3	Total error	Verbal WM	-0.049	-0.239	-2.048	0.044
	Operand-unrelated distractor error	Verbal STM	0.095	0.408	3.603	0.001
		Verbal WM	-0.069	-0.242	-2.132	0.037
4	Total error	Visuo-spatial WM	-0.041	-0.268	-2.308	0.024

learnt. The suggestion of the interacting neighbors model even holds for those young children in grades 3 and 4. The model assumes that the operand-related distractors lead to stronger confounding with the solutions than the operand-unrelated distractors.

However, as regards the operand-relatedness effect, we found no difference between grades 3 and 4. In fact, there was an operand-relatedness effect in both grades but it was neither stronger nor weaker than in the other grade. This result was again in line with the only longitudinal study of multiplication in a verification paradigm in children (De Brauwer and Fias, 2009). The finding of the present study is consistent with the idea that multiple changes may occur in the associative network. First, the strength of the association network increases with age and experience (which leads to faster retrieval in older children). Second, the network may become more refined in reciprocal inhibition. More association strength with age would lead to a higher operand relatedness effect because related entries are activated more. However, better reciprocal inhibition would lead to better differentiation between entries and therefore to a lower operand relatedness effect because related entries could be more easily inhibited. If both processes increase similarly with age and experience, the operand-relatedness effect may stay unchanged. This is what we found in the present study.

### **An Age-Related Shift from Verbal to Visuo-Spatial Working Memory Predicting Multiplication Performance**

Interestingly, we found that verbal WM predicts multiplication problem solving in grade 3, while in grade 4 visuo-spatial WM is the predictor. This finding for multiplication performance extends and refines current accounts of the role of different WM components during different developing stages. A developmental change of the influence of verbal and visuo-spatial components was reported several years ago for more general math capabilities: it was shown that there is a strong link between verbal and mathematical skills when young children are learning new information which becomes weaker in older children as the result of practice (Jensen, 1980). In accordance to this finding, several studies have shown the weak conjunction between phonological loop and mathematical performance in adults (e.g., Logie and Baddeley, 1987; Heathcote, 1994; Logie et al., 1994). The present study did not find any significant correlation between verbal WM and multiplication performance in grade 4 which can be related to a gradual shift from strongly verbal representations of multiplication to the build-up of a more abstract semantic retrieval of mathematical facts from long-term memory which is visually based, at least when the stimuli are presented visually as in our study.

One possible suggestion is that one may expect to see more predictability of verbal WM in grade 4. However, this was not the case. Three reasons may explain this finding. First, learning and task context of multiplication problems encountered in (Austrian) schools may contribute to their explanation. While in the initial learning phase in grades 2 and 3, multiplication problems may be more auditorily and verbally trained, they may

be more often encountered visually as part of more complex arithmetic problems in grade 4. Second, the shift toward more visuo-spatial processing is consistent with previous studies on arithmetic development showing that in children, arithmetic tasks require superior demand of visuo-spatial processing during the development (Alloway and Passolunghi, 2011). In fact in adults, Fürst and Hitch (2000) showed that the phonological loop is not crucially caught up in retrieving factual mathematical knowledge which is also consistent with our data that verbal WM plays a lesser role in older children. Finally, the same verbal to visuo-spatial WM shift has been observed in other arithmetic domains. Meyer et al. (2010) found such a shift from grade 2 to grade 3 in some basic arithmetic and mathematical reasoning. For these reasons, we believe that our finding of a developmental shift from verbal to visuo-spatial WM with age and experience does not come as a surprise but is actually consistent with the literature in other fields of arithmetic development. In sum, the data shows an important developmental shift from verbal to visuo-spatial WM in the prediction of simple multiplication problem performance (as indexed by overall errors) from grade 3 to grade 4.

Furthermore, neuroimaging studies revealed a neural dissociation of verbal and visuo-spatial WM (Smith et al., 1996; Thürling et al., 2012), which were modified differently due to arithmetic training. The brain activation pattern of development and training of calculation shows a shift of activation from the frontal to the parietal regions (for a review Zamarian et al., 2009). This modification shows a shift from verbally representation of the calculation to more visually representation. While the frontal area is involved in verbal WM, the parietal area is mostly involved in visuo-spatial WM (for a review Cabeza and Nyberg, 2000; Dumontheil and Klingberg, 2012).

Interestingly, for the operand-unrelated distractor errors in grade 3, verbal STM reached significance as the only STM predictor in our whole study. However, this makes sense because during the second and third years of elementary school children are commonly highly trained with direct verbal learning of multiplication facts. Therefore, verbal STM is still significant for multiplication in grade 3. In the fourth grade, however, children have to use the learnt skills, such as multiplication, indirectly in more advanced mathematic problems such as mathematical text questions which does not involve any aspect of STM massively in this grade. Verbal STM may only affect the operand-unrelated distractor errors because the operand-relatedness may lead to interference specifically in the STM where no information is manipulated. Vice versa, the solutions share at least one element with possible operand-related distractors. It seems plausible that in such clear cases which require no manipulation and selection of information, verbal STM processing is most predictive. Again, our finding that verbal STM influences multiplication performance in earlier grades is consistent with previous findings from other more general arithmetic measures. For instance, Alloway and Passolunghi (2011) showed that verbal and visuo-spatial STM were involved in arithmetic performance at age 7 but only visuo-spatial STM was involved at age 8. Although the prediction of operand-unrelated distractor error by verbal STM in grade 3 was reasonable, the positive correlation between verbal

STM and operand-unrelated distractor error was unexpected. One possible explanation would be the interference of other simultaneous processes, which occupy STM. We know that the results of simple multiplication problems are retrieved from long-term memory (for a review of neuroimaging studies see Zamarian et al., 2009). Indeed, the results of the one-digit time one-digit multiplication problems, which belong to the multiplication table are stored in long-term memory and retrieved via WM. Therefore, it may be concluded that to answer these problems, we do not rely so much on STM (Butterworth et al., 1996). Hence, any involvement of STM in other simultaneous processes can interfere with this fact retrieval procedure. But this is not the case of WM. We know that WM is involved in almost every cognitive process. Since WM has a crucial role in the retrieving of multiplication result, higher WM capacity can lead to a better manipulation on different processing including multiplication performance. Butterworth et al. (1996) showed that in a patient with impaired STM, the mental calculations such as one-digit multiplication are intact. However, we believe that this is only a possible interpretation, which needs to be tested directly.

None of the memory components were able to predict RTs in both grades. We believe that this is due to high (inter-individual and intra-individual) variability in the RT measures for the children, which may be overcome in comparisons of means but may be critical for inter-individual comparisons and correlations. Variability in RTs can be explained by several sources. First, children use different strategies for multiplication problem solving (Cooney et al., 1988; Sherin and Fuson, 2005) which mostly lead to equal (correct) responses but to different RTs. Second, individual differences in mathematical competence modulate RTs during mental arithmetic. For instance, Grabner et al. (2007) suggested that the recruitment of retrieval strategies during arithmetic problem solving may be caused by individual differences in mathematical ability. Therefore, different children rely on different memory processes. This may lead to highly variable RTs, not only intra-individually, but also inter-individually, even though both ways may lead to the solution of the multiplication problem. For these reasons, RT may be more sensitive to intra- and inter-individual variability than errors. Future studies should probably combine investigations of the strategy used and different WM components to examine if specific WM components are associated with specific solving strategies.

## Conclusion

In line with the previous findings (Swanson, 2006; Meyer et al., 2010), the current study suggests that although verbal WM may facilitate early stages of arithmetic learning and performance, visuo-spatial WM may support later arithmetic performance during the development – at least during elementary school. We would like to mention that while we found this shift in prediction of multiplication problem solving from grade 3 to 4, the others found it in different ages, however, albeit for different mathematical contents. For instance Meyer et al. (2010) found the shift in mathematical reasoning from grade 2 to 3.

Meyer et al. (2010) were concerned with mathematical reasoning. Their mathematical reasoning subtest of the WIAT-II “is a verbal problem solving test that measures the ability to count, identify geometric shapes and solve single- and multi-step word problems.” In contrast, we were concerned with multiplication. Multiplication – as said above – is introduced in grade 2, verbally trained in grade 3 and then integrated in visual tasks in grade 4 – therefore the shift from verbal to visual makes sense for multiplication at exactly that age. Because the mathematical reasoning subtest of the WIAT-II is an aggregate score of many different tasks, it is hard to tell, why the shift was caused in Meyer et al. (2010) from grade 2 to 3. However, because the subtests contained some very basic tasks like counting or identifying geometric shapes, which are introduced earlier than multiplication, it is possible that the shift from verbal to visuo-spatial WM is also earlier in their study. In sum, it seems that this shift may be found in different developing ages for differing mathematical skills. This shift may serve as an essential step in mathematical development, however, its relation to age may vary according to mathematical content – in our view, this deserves further more detailed investigation in the future.

This changing role of verbal and visuo-spatial WM components for predicting arithmetic performance could be useful for diagnosis and intervention in children with mathematical learning difficulties. However, we recommend that future studies should also assess children’s strategy-use. By examining strategy-use together with the contribution of different memory components, researchers might be able to uncover cognitive demands of multiplication learning in developmental ages.

As regards the fact retrieval network itself, the current data suggest that retrieval is faster and more efficient from grade 3 to grade 4; however, the lack of change in the operand-relatedness effect with age may suggest that in children’s fact retrieval network both automatic association and reciprocal inhibition of concurrent responses may increase. More associations and at the same time better inhibition might lead to an unaltered operand-relatedness effect in this longitudinal study. This is only a speculative interpretation which needs to be examined in future studies with considering inhibitory control, attentional processing, and self-regulation as well.

## Acknowledgments

We would like to thank all participating children and their parents as well as the involved school teachers and principals. This research was funded by a grant from the AKTION Österreich- Tschechien (45p13) to SP and HCN, as well as by a grant from the Science Campus Tübingen, project 8.4 to HCN supporting MS. HCN’s research is further supported by the LEAD Graduate School [GSC1028], funded by the Excellence Initiative of the German federal and state governments. We acknowledge support by Deutsche Forschungsgemeinschaft and Open Access Publishing Fund of University of Tübingen. Finally, we thank Amanda Lillywhite and Jennifer Proehl for the proofreading of the manuscript.

## References

- Alloway, T. P., and Alloway, R. G. (2010). Investigating the predictive roles of working memory and IQ in academic attainment. *J. Exp. Child Psychol.* 106, 20–29. doi: 10.1016/j.jecp.2009.11.003
- Alloway, T. P., Gathercole, S. E., and Pickering, S. J. (2006). Verbal and visuospatial short-term and working memory in children: are they separable? *Child Dev.* 77, 1698–1716. doi: 10.1111/j.1467-8624.2006.00968.x
- Alloway, T. P., and Passolunghi, M. C. (2011). The relationship between working memory, IQ, and mathematical skills in children. *Learn. Individ. Differ.* 21, 133–137. doi: 10.1016/j.lindif.2010.09.013
- Ashcraft, M. H. (1987). “Children’s knowledge of simple arithmetic: a developmental model and simulation,” in *Formal Methods in Developmental Psychology*, (New York, NY: Springer), 302–338.
- Baddeley, A. (1992). Working memory. *Science* 255, 556–559. doi: 10.1126/science.1736359
- Baddeley, A. (1998). Random generation and the executive control of working memory. *Q. J. Exp. Psychol. A* 51, 819–852. doi: 10.1080/713755788
- Baddeley, A. D., and Hitch, G. (1974). Working memory. *Psychol. Learn. Motiv.* 8, 47–89. doi: 10.1016/S0079-7421(08)60452-1
- Butterworth, B., Cipolotti, L., and Warrington, E. K. (1996). Short-term memory impairment and arithmetical ability. *Q. J. Exp. Psychol. A* 49, 251–262. doi: 10.1080/027249896392892
- Butterworth, B., Marchesini, N., and Girelli, L. (2003). “Basic multiplication combinations: passive storage or dynamic reorganization?” in *The Development of Arithmetic Concepts and Skills: Constructive Adaptive Expertise*, eds A. J. Baroody and Dowker (Mahwah: LEA), 187.
- Cabeza, R., and Nyberg, L. (2000). Imaging cognition II: an empirical review of 275 PET and fMRI studies. *J. Cogn. Neurosci.* 12, 1–47. doi: 10.1162/08989290051137585
- Campbell, J. I. (1995). Mechanisms of simple addition and multiplication: A modified network-interference theory and simulation. *Math. Cogn.* 1, 121–164.
- Campbell, J. I. (1997). On the relation between skilled performance of simple division and multiplication. *J. Exp. Psychol. Learn. Mem. Cogn.* 23, 1140–1159. doi: 10.1037/0278-7393.23.5.1140
- Campbell, J. I., and Graham, D. J. (1985). Mental multiplication skill: Structure, process, and acquisition. *Can. J. Psychol.* 39, 338–366. doi: 10.1037/h0080065
- Cooney, J. B., Swanson, H. L., and Ladd, S. F. (1988). Acquisition of mental multiplication skill: evidence for the transition between counting and retrieval strategies. *Cogn. Instr.* 5, 323–345. doi: 10.1207/s1532690xci0504\_5
- Corsi, P. M. (1973). *Human Memory and the Medial Temporal Region of the Brain*. Ph.D. thesis, Department of Psychology, ProQuest Information & Learning, Ann Arbor.
- Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychol. Bull.* 104, 163–191. doi: 10.1037/0033-2909.104.2.163
- Cowan, N. (2008). What are the differences between long-term, short-term, and working memory? *Prog. Brain Res.* 169, 323–338. doi: 10.1016/S0079-6123(07)00020-9
- De Brauwier, J., and Fias, W. (2009). A longitudinal study of children’s performance on simple multiplication and division problems. *Dev. Psychol.* 45, 1480–1496. doi: 10.1037/a0015465
- Domahs, F., Delazer, M., and Nuerk, H.-C. (2006). What makes multiplication facts difficult: problem size or neighborhood consistency? *Exp. Psychol.* 53, 275–282. doi: 10.1027/1618-3169.53.4.275
- Dumontheil, I., and Klingberg, T. (2012). Brain activity during a visuospatial working memory task predicts arithmetical performance 2 years later. *Cereb. Cortex* 22, 1078–1085. doi: 10.1093/cercor/bhr175
- Friedman, N. P., and Miyake, A. (2000). Differential roles for visuospatial and verbal working memory in situation model construction. *J. Exp. Psychol. Gen.* 129, 61–83. doi: 10.1037/0096-3445.129.1.61
- Fürst, A. J., and Hitch, G. J. (2000). Separate roles for executive and phonological components of working memory in mental arithmetic. *Mem. Cognit.* 28, 774–782. doi: 10.3758/BF03198412
- Geary, D. C., Bow-Thomas, C. C., Liu, F., and Siegler, R. S. (1996). Development of arithmetical competencies in Chinese and American children: influence of age, language, and schooling. *Child Dev.* 67, 2022–2044. doi: 10.2307/1131607
- Grabner, R. H., Ansari, D., Reishofer, G., Stern, E., Ebner, F., and Neuper, C. (2007). Individual differences in mathematical competence predict parietal brain activation during mental calculation. *Neuroimage* 38, 346–356. doi: 10.1016/j.neuroimage.2007.07.041
- Heathcote, D. (1994). The role of visuo-spatial working memory in the mental addition of multi-digit addends. *Cah. Psychol. Cogn. Curr. Psychol. Cogn.* 13, 207–245.
- Hitch, G. J., Halliday, S., Schaafstal, A. M., and Schraagen, J. M. C. (1988). Visual working memory in young children. *Mem. Cognit.* 16, 120–132. doi: 10.3758/BF03213479
- Isaacs, E. B., and Vargha-Khadem, F. (1989). Differential course of development of spatial and verbal memory span: a normative study. *Br. J. Dev. Psychol.* 7, 377–380. doi: 10.1111/j.2044-835X.1989.tb00814.x
- Jarvis, H. L., and Gathercole, S. E. (2003). Verbal and non-verbal working memory and achievements on national curriculum tests at 11 and 14 years of age. *Educ. Child Psychol.* 20, 123–140.
- Jensen, A. R. (1980). *Bias in Mental Testing*. Florence: Free Press.
- Koshmider, J. W., and Ashcraft, M. H. (1991). The development of children’s mental multiplication skills. *J. Exp. Child Psychol.* 51, 53–89. doi: 10.1016/0022-0965(91)90077-6
- Lemaire, P., Abdi, H., and Fayol, M. (1996). The role of working memory resources in simple cognitive arithmetic. *Eur. J. Cogn. Psychol.* 8, 73–104. doi: 10.1080/095414496383211
- Lemaire, P., and Siegler, R. S. (1995). Four aspects of strategic change: contributions to children’s learning of multiplication. *J. Exp. Psychol. Gen.* 124, 83–97. doi: 10.1037/0096-3445.124.1.83
- Logie, R. H., and Baddeley, A. D. (1987). Cognitive processes in counting. *J. Exp. Psychol. Learn. Mem. Cogn.* 13, 310–326. doi: 10.1037/0278-7393.13.2.310
- Logie, R. H., Gilhooly, K. J., and Wynn, V. (1994). Counting on working memory in arithmetic problem solving. *Mem. Cognit.* 22, 395–410. doi: 10.3758/BF03200866
- Masson, M. E. (2011). A tutorial on a practical Bayesian alternative to null-hypothesis significance testing. *Behav. Res.* 43, 679–690. doi: 10.3758/s13428-010-0049-5
- McKenzie, B., Bull, R., and Gray, C. (2003). The effects of phonological and visual-spatial interference on children’s arithmetical performance. *Educ. Child Psychol.* 20, 93–108.
- Meyer, M., Salimpoor, V., Wu, S., Geary, D., and Menon, V. (2010). Differential contribution of specific working memory components to mathematics achievement in 2nd and 3rd graders. *Learn. Individ. Differ.* 20, 101–109. doi: 10.1016/j.lindif.2009.08.004
- Miyake, A., Friedman, N. P., Rettinger, D. A., Shah, P., and Hegarty, M. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related? A latent-variable analysis. *J. Exp. Psychol. Gen.* 130, 621–640. doi: 10.1037/0096-3445.130.4.621
- Miyake, A., and Shah, P. (1999). *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139174909
- Nuerk, H.-C., Kaufmann, L., Zoppoth, S., and Willmes, K. (2004). On the development of the mental number line: more, less, or never holistic with increasing age? *Dev. Psychol.* 40, 1199–1211. doi: 10.1037/0012-1649.40.6.1199
- Raftery, A. E. (1995). Bayesian model selection in social research. *Sociol. Methodol.* 25, 111–164. doi: 10.2307/271063
- Rasmussen, C., and Bisanz, J. (2005). Representation and working memory in early arithmetic. *J. Exp. Child Psychol.* 91, 137–157. doi: 10.1016/j.jecp.2005.01.004
- Shah, P., and Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: an individual differences approach. *J. Exp. Psychol. Gen.* 125, 4–27. doi: 10.1037/0096-3445.125.1.4
- Sherin, B., and Fuson, K. (2005). Multiplication strategies and the appropriation of computational resources. *J. Res. Math. Educ.* 36, 347–395.
- Simmons, F., Singleton, C., and Horne, J. (2008). Brief report Phonological awareness and visual-spatial sketchpad functioning predict early arithmetic attainment: evidence from a longitudinal study. *Eur. J. Cogn. Psychol.* 20, 711–722. doi: 10.1080/09541440701614922
- Smith, E. E., Jonides, J., and Koeppel, R. A. (1996). Dissociating verbal and spatial working memory using PET. *Cereb. Cortex* 6, 11–20. doi: 10.1093/cercor/6.1.11

- Swanson, H. L. (2006). Cross-sectional and incremental changes in working memory and mathematical problem solving. *J. Educ. Psychol.* 98, 265–281. doi: 10.1037/0022-0663.98.2.265
- Thüring, M., Hautzel, H., Küper, M., Stefanescu, M., Maderwald, S., Ladd, M. E., et al. (2012). Involvement of the cerebellar cortex and nuclei in verbal and visuospatial working memory: a 7T fMRI study. *Neuroimage* 62, 1537–1550. doi: 10.1016/j.neuroimage.2012.05.037
- Vandierendonck, A., Kemps, E., Fastame, M. C., and Szmalec, A. (2004). Working memory components of the Corsi blocks task. *Br. J. Psychol.* 95, 57–79. doi: 10.1348/000712604322779460
- Verguts, T., and Fias, W. (2005). Interacting neighbors: a connectionist model of retrieval in single-digit multiplication. *Mem. Cognit.* 33, 1–16. doi: 10.3758/BF03195293
- Winer, B. J., Brown, D. R., and Michels, K. M. (1971). *Statistical Principles in Experimental Design*. New York, NY: McGraw-Hill.
- Zamarian, L., Ischebeck, A., and Delazer, M. (2009). Neuroscience of learning arithmetic—evidence from brain imaging studies. *Neurosci. Biobehav. Rev.* 33, 909–925. doi: 10.1016/j.neubiorev.2009.03.005

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Soltanlou, Pixner and Nuerk. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

