UNIVERSITY *of* York

This is a repository copy of *Twenty-four or Four-and-twenty:Language Modulates Cross-Modal Matching for Multi-Digit Numbers in Children and Adults*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/167189/

Version: Published Version

White Rose
university consortium
Universities of Leeds, Sheffield & York

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Twenty-four or four-and-twenty: Language modulates cross-modal matching for multidigit numbers in children and adults

Anna F. Steiner [a,*], Chiara Banfi [a], Sabrina Finke [a], Ferenc Kemény [a], Francina J. Clayton [b], Silke M. Göbel [b,c], Karin Landerl [a,d]

[a] Institute of Psychology, University of Graz, 8010 Graz, Austria
[b] Department of Psychology, University of York, Heslington, York YO10 5DD, UK
[c] Department of Special Needs Education, University of Oslo, 0316 Oslo, Norway
[d] Department of Cognitive Science, Macquarie University, Sydney, New South Wales 2109, Australia

ARTICLE INFO

ABSTRACT

Does number–word structure have a long-lasting impact on transcoding? Contrary to English, German number words comprise decade–unit inversion (e.g., *vierundzwanzig* is literally translated as *four-and-twenty*). To investigate the mental representation of numbers, we tested the effect of visual and linguistic–morphological characteristics on the development of verbal–visual transcoding. In a longitudinal cross-linguistic design, response times (RTs) in a number-matching experiment were analyzed in Grade 2 (119 German-speaking and 179 English-speaking children) and in Grade 3 (131 German-speaking and 160 English-speaking children). To test for long-term effects, the same experiment was given to 38 German-speaking and 42 English-speaking adults. Participants needed to decide whether a spoken number matched a subsequent visual Arabic number. Systematic variation of digits in the nonmatching distractors allowed comparison of three different transcoding accounts (lexicalization, visual, and linguistic–morphological). German speakers were generally slower in rejecting inverted number distractors than English speakers. Across age groups, German speakers were more distracted by Arabic numbers that included the correct unit digit, whereas English speakers showed stronger distraction when the correct decade digit was included. These RT patterns reflect differences in number–word

* Corresponding author.
    E-mail address: a.steiner@uni-graz.at (A.F. Steiner).

morphology. The individual cost of rejecting an inverted distractor (inversion effect) predicted arithmetic skills in German-speaking second-graders only. The moderate relationship between the efficiency to identify a matching number and arithmetic performance could be observed cross-linguistically in all age groups but was not significant in German-speaking adults. Thus, findings provide consistent evidence of a persistent impact of number–word structure on number processing, whereas the relationship with arithmetic performance was particularly pronounced in young children.

Introduction

When we see the Arabic number *24,* we cannot help but coactivate the corresponding number word *twenty-four.* However, it is still a matter of debate how we shift from one symbolic number representation to another and how this cross-modal shifting develops during childhood, especially in the case of multidigit numbers. In Dehaene's (1992) triple-code model, the seemingly effortless switching between number formats is explained in terms of transfer among three interlinked numerical codes. The auditory–verbal word frame relies on a general-purpose language module and manipulates spoken number words, the visual Arabic number form processes numbers in Arabic notation, and the analogue magnitude code represents number semantics as abstract numerical quantities. Although the Arabic number system is used in most languages today (Chrisomalis, 2010), number–word structure is language dependent and not always consistent with Arabic multidigits. An important and common discrepancy is the inversion rule (e.g., German, Dutch, Slovenian), requiring to reverse the order of decades and units (e.g., *24* is literally pronounced as *four-and-twenty*).

The current study investigated the effect of morphological word structure on the development of transcoding procedures by comparing German, a language with inversion, with English (no inversion). In particular, we analyzed response time (RT) patterns for distractor types with varying degrees of overlap in a verbal–visual number-matching paradigm (e.g., target: *twenty-four;* distractors: *42, 28, 48,* and *36*) in order to specifically test different theoretical accounts of number representation. Because we were particularly interested in the development of integrating tens and units into two-digit numbers, when the required basic procedures were already learned, we compared task performance at three different levels of transcoding expertise (Grade 2, Grade 3, and adults) in both languages.

Theoretical models of transcoding differ in whether or not accessing number semantics is an obligatory component of the process. Semantic models (e.g., McCloskey, 1992; Power & Dal Martello, 1990) are based on the assumption that the verbal (or visual Arabic) input is first transformed into an abstract internal magnitude representation and subsequently transformed into the transcoded output (e.g., *twenty-four* →2 tens, 4 units →*24*). Asemantic models allow direct cross-modal transfer between the two symbolic number forms without accessing magnitude representations (e.g., Barrouillet, Camos, Perruchet, & Seron, 2004; Dotan & Friedmann, 2018; Power & Dal Martello, 1997). To date, only the ADAPT model (A Developmental, Asemantic, and Procedural model for Transcoding; Barrouillet et al., 2004) attempts to explain transcoding development. In this model, transcoding entails lexical components as well as procedural components. Number writing starts with parsing the phonological input within the verbal word frame, which is temporarily stored in a phonological buffer. A parsing process identifies each single word unit. As soon as a chunk is found that corresponds with lexical units in long-term memory, the process stops and the next input chunk is analyzed. Long-term memory stores lexical primitives, including the early-acquired units (1–9), later on complemented by teens, decades, and separators such as *hundred* and *thousand.* Improvement in transcoding is explained by continuous development of the procedural system as well as an increasing number of representational units stored in long-term memory. Accordingly, Barrouillet et al. (2004)

claimed that, in the (noninverted) French language, children have incorporated two-digit numbers in their mental lexicon as early as Grade 2. Thus, instead of relying on procedures, the entire number is directly retrieved from long-term memory.

Contrary to the assumption of progressive lexicalization of multidigit numbers, there is some evidence that morphological characteristics of number words may have a lifelong impact. In an English–French comparison (both noninverted languages), 10-year-olds with ample experience with double-digit numbers showed differences in an RT-based transcoding task; RTs for numbers above 60 were longer in French than in English, probably because of the partly vigesimal number–word system (*eighty* is literally spoken as *four-twenty*) that might hamper lexicalization. The two language groups did not differ in RTs for numbers with comparable number–word structures (Van Rinsveld & Schiltz, 2016). However, it is unclear whether these results can easily be generalized to inversion, where number names are unambiguous and only decade–unit order is reversed.

It comes as no surprise that studies found early problems with transcoding multidigit numbers in languages with inversion property. For instance, German-speaking 7-year-olds had more difficulties with digit order in number writing than Japanese-speaking peers acquiring a highly transparent number–word system (Moeller, Zuber, Olsen, Nuerk & Willmes, 2015). Similarly, Dutch-speaking second-graders made more inversion errors than children speaking French, a language without inversion (Imbo, Vanden Bulcke, De Brauwer, & Fias, 2014). In line with the assumption that decade–unit inversion affects mostly young learners, van der Ven, Klaiber, and van der Maas (2017) reported a sharp decrease in the number of inversion errors in Dutch between kindergarten and Grade 3. However, even though such errors were quite exceptional later on, they did occur up to Grade 6, indicating an impact of inversion during later development. That being said, overall accuracy rates in conventional transcoding tasks are typically close to ceiling for older children and adults, and this might obscure enduring language differences in number reading and number writing (Zuber, Pixner, Moeller, & Nuerk, 2009). RT-based paradigms may help to reveal such differences, even if responses are correct.

Such an RT-based transcoding paradigm was recently used by Poncin, Van Rinsveld, and Schiltz (2019). Participants heard a number word first and needed to choose the corresponding Arabic number from four options. The constituent digits of the Arabic number were presented either simultaneously or with a 500-ms delay of either the decade or the unit. Interestingly, slower responses in inverted German than in noninverted French for simultaneous and tens-first conditions could still be observed among fourth-graders but not among adults. Indirect evidence of an impact of inversion even in adults was only reported in a comparison task with double digits (Moeller, Shaki, Göbel, & Nuerk, 2015; Nuerk, Weger, & Willmes, 2005). For number pairs in which the magnitude decision for units only was incompatible with the magnitude decision for decades (and therefore the full number; e.g., 46_72 for which 4 < 7 but 6 > 2), RTs were higher for German speakers than for English speakers, perhaps because the unit is more prominent (it comes first) in number words. In summary, although previous evidence is suggestive of persistent effects of number–word morphology on transcoding, no study has as yet reported such effects across different ages.

It is also as yet unclear to what extent visual overlap between digits provides an alternative explanation of RT patterns in cross-modal number matching. Such a visual account is based on the assumption that spoken number words might be mentally transformed into Arabic numbers, followed by a comparison of visual features shared between this mental imagery and the visually presented numbers. This account was first proposed by Cohen (2009), who found an impact of visual similarity in an audio–visual matching task with single-digit numbers (Cohen, Warren, & Blanc-Goldhammer, 2013). It might also be fruitful to investigate whether the presence of matching digits affects RT patterns for two-digit numbers; nonmatching distractors can be created so that one digit is visually identical to the target and in the same position (e.g., *twenty-four → 28*) or is visually identical but in a different position (e.g., *twenty-four → 82*).

The high relevance of multidigit transcoding is evident from previous studies showing associations with arithmetic achievement in languages with number–word inversion (Imbo et al., 2014; Moeller, Pixner, Zuber, Kaufmann, & Nuerk, 2011; van der Ven et al., 2017) as well as without number–word inversion (Göbel, Watson, Lervåg, & Hulme, 2014; Simmons, Willis, & Adams, 2012). Interestingly, Göbel, Moeller, Pixner, Kaufmann, and Nuerk (2014) found that additions with carry operations

required more processing time among German-speaking second-graders than among Italian-speaking second-graders, indicating that number–word structure, and in particular number–word inversion, may have a direct impact on arithmetic performance.

The current study

Previous studies on multidigit transcoding in children have mostly investigated errors during number writing (Imbo et al., 2014; Moeller, Huber, Nuerk, & Willmes, 2011; Pixner et al., 2011; van der Ven et al., 2017; Van Rinsveld & Schiltz, 2016; Zuber et al., 2009) or reading (Dowker, Bala, & Lloyd, 2008; Moura et al., 2013). However, over time errors become exceptional for two-digit numbers. To investigate correct transcoding of two-digit numbers beyond early development, we applied an RT-based number-matching paradigm. The impact of number–word inversion was investigated in a cross-linguistic design comparing samples of German- and English-speaking children and adults. English and German are ideal for such a comparison because they are closely related Germanic languages with highly similar lexical number primitives (e.g., *drei–three, nineteen–neunzehn, fünfzig–fifty, hundert–hundred*), but only German applies decade–unit inversion (e.g., *vierundzwanzig–twenty-four*).

In the number-matching task, participants first heard a number word and then were shown an Arabic digit string. The two numbers either matched (e.g., *twenty-four → 24*) or did not match (e.g., *twenty-four → 28*). Systematic variation of position and identity of digits in the nonmatching distractors allowed us to investigate the influence of visual and linguistic–morphological number characteristics on cross-modal matching. The following distractor types were used:

1. Inverted numbers fully preserving digit identity, but with inverted digit position compared with the preceding number word: The decade (D) digit was in the unit (U) position and vice versa (e.g., *twenty-four → 42*; U+D+).
2. The decade was preserved in position and identity, whereas the unit was different (e.g., *twenty-four → 28;* D+U−).
3. The unit was preserved in identity but presented in decade position. The second digit had no overlap (e.g., *twenty-four → 48;* U+D−).
4. The unit was preserved in position and identity, whereas the decade was different[1] (e.g., *twenty-four → 84;* D−U+).
5. Nonrelated distractors (e.g., *twenty-four → 36;* D−U−).

Based on participants' difficulties with rejecting each of these distractors, we aimed to test three competing hypotheses. The lexicalization account is based on the proposition that two-digit numbers are lexicalized early in development (Barrouillet et al., 2004). Thus, we would expect no major differences between distractor conditions because they all deviated from the correct lexical entry.

A visual account would predict that Arabic digits overlapping with the target in identity should negatively influence the speed of rejection. Maximum interference would be expected for U+D+ distractors because digit identity is fully preserved and only digit position is inverted. Distractors with overlap in either the unit or the decade (e.g., *twenty-four: 28* D+U−, *48* U+D−, *84* D−U+) would induce intermediate interference, with perhaps somewhat more difficulty to reject D+U− (and D−U+) than U+D− because the latter one retains identity but not position of the unit number. The control condition with no visual overlap should again be easier to reject than all other distractors. Importantly, this account would not predict any marked differences between German and English because the visual overlap is obviously the same.

A linguistic–morphological account would predict language-specific RT patterns. Based on the procedural transcoding rules for multidigit numbers in the ADAPT model (Barrouillet et al., 2004), we would expect that morphological number–word structure influences task performance. We would once again predict most interference if both constituent digits are included in the inverted Arabic number (U+D+). Critically, however, we would expect that the inversion effect is larger in German

---

[1] This distractor type was used only in children's Grade 3 assessments.

than in English if the inverted Arabic number is processed in standard reading direction (left to right) because it shows stronger overlap with the verbal number form.

Language-dependent differences would also be expected for distractors that have either the decade or the unit of the prior number word in the left-most position. In English, U+D− distractors should be easier to reject than D+U− distractors because the left digit is not the one that would be expected when hearing the start of the number word (e.g., *twenty-four → 48*). For German, matching is a bit more complicated. If we assume that Arabic numbers are processed in the standard reading direction, we would expect more interference for U+D− than for D+U− (and D−U+) because the U+D− distractor starts with the same number as the number word (e.g., *four-and-twenty → 48*).

It is, however, also possible that German-speaking participants adapt the processing direction toward the inverted number–word structure and process Arabic numbers in a right-to-left fashion. There is indeed some evidence for such stimulus-driven adaptations in number processing. Huber, Mann, Nuerk, and Moeller (2014) observed that in a comparison task with double digits, German-speaking adults shifted their eye fixations between units and tens depending on whether items could be mostly solved based on decade comparison or unit comparison. Similarly, a sample of 10 Dutch-speaking 9-year-olds showed alternating fixation strategies at units and tens in a number line task. Fixation patterns included (a) equal fixations at both digits, indicating holistic processing, (b) fixating first the tens and then briefly the unit revealing a decomposed parallel processing, and (c) decomposed sequential fixations with focusing on the tens and almost not at the units (van Viersen, Slot, Kroesbergen, van't Noordende, & Leseman, 2013). Note that these studies did not provide evidence for a clearly inverted processing direction in the inverted languages German and Dutch. Still, if in our matching paradigm German-speaking participants indeed process numbers in an adapted right-to-left fashion, both the D+U− distractor (e.g., *twenty-four → 28*) and the U+D− distractor (e.g., *twenty-four → 48*) should be easy to reject because neither of them has the correct unit on the right-hand side. The D−U+ distractor (*four-and-twenty → 84;* included in Grade 3 assessments only) should be relatively hard to reject.

In summary, our study goes beyond earlier research in several ways: First, we investigated RT patterns of correct audio–visual transcoding and not only when the transcoding process goes wrong as in the case of errors during number reading or writing. Second, this is the first direct cross-linguistic investigation of the effect of inversion in two highly similar languages (German and English). Third, we went beyond analyzing the cost of items with decade–unit inversion by presenting additional distractor conditions in which we systematically varied identity and position of the unit or the decade number in order to better understand how two-digit numbers are represented in the cognitive system. Furthermore, we investigated the developmental challenges of cross-modal number processing longitudinally across the critical period from Grade 2 to Grade 3, when children are practicing arithmetic operations with two-digit numbers and thus get ample practice in reading and writing such numbers. We were also able to compare children's performance patterns cross-sectionally with adult samples in both languages.

Last but not least, we aimed to contribute further evidence on the association of cross-modal number matching with concurrent arithmetic performance in children and adults. In line with previous research using conventional transcoding tasks such as writing and identifying multidigit numbers, we expected a positive relationship of matching efficiency with arithmetic in both a language with inversion (Imbo et al., 2014; Moeller et al., 2011; van der Ven et al., 2017) and a language without inversion (Göbel, Watosn, et al., 2014; Simmons et al., 2012). Studies investigating the relationship between audio–visual matching and arithmetic performance are typically limited to single digits, and results have been inconsistent (Lin & Göbel, 2019; Lyons, Price, Vaessen, Blomert, & Ansari, 2014; Sasanguie, Lyons, De Smedt, & Reynvoet, 2017; Sasanguie & Reynvoet, 2014).

The experimental nature of the number-matching paradigm also allowed us to explore whether a relation of the specific cost to reject an inverted distractor to arithmetic performance would be evident. If so, we suspect that it might be stronger in German; because units and tens occur in different order depending on number format, it might be more relevant to suppress incorrect digit order when doing arithmetic. The focus of the study was on investigating number matching in children. The task was administered as part of a larger longitudinal study in which we also measured established domain-general predictors of arithmetic. Hence, we could test whether transcoding would account

for variance in arithmetic performance above and beyond age, intelligence, and working memory, in line with previous studies with English-speaking samples (e.g., Göbel, Watosn, et al., 2014; Reeve, Reynolds, Paul, & Butterworth, 2018).

Method

Participants

Children were recruited from 5 primary schools in Graz (Austria) and 11 primary schools across Yorkshire (United Kingdom). Overall, 148 monolingual German-speaking children (67 female) and 245 monolingual English-speaking children (116 female) participated at the end of Grade 2. Groups were matched on duration of formal education. Because schooling starts 1 year earlier in the United Kingdom than in Austria, German-speaking children were on average 11 months older than their English-speaking counterparts (German mean age = 8;2 [years;months], *SD* = 0;4; English mean age = 7;3, *SD* = 0;4). One year later in Grade 3, we were able to reassess 145 German-speaking children (65 female; mean age = 9;1, *SD* = 0;4) and 190[2] English-speaking children (89 female; mean age = 8;3, *SD* = 0;4).

To check whether there is a consistent pattern in the number-matching experiment across ages, we also recruited an adult sample consisting of 38 native German-speaking participants (26 female; mean age = 23;6, *SD* = 3;1) and 42 native English-speaking participants (30 female; mean age = 19;8, *SD* = 2;2). Adults received only the number-matching task and an assessment of arithmetic fluency (see below). No further predictors were assessed, so that a smaller sample size was sufficient.

This study was performed in accordance with the latest version of the Declaration of Helsinki and in compliance with national legislation. The University of Graz and the University of York Psychology Department ethics committees approved the study, and written informed consent was obtained from participants or their legal guardians.

Materials and procedure

Children carried out the number-matching experiment in the context of an extensive task battery with individually administered tasks (number matching and working memory) as well as class-administered tasks (nonverbal reasoning and arithmetic) within the scope of a longitudinal project. Adults performed only the number-matching experiment and an arithmetic fluency task. Individual assessments were administered in a separate room at the school (children) or university (adults).

Number matching

This task was presented in PsychoPy Version 1.85.3 (Peirce, 2007). Children sat in front of a laptop, and adults sat in front of a desktop computer or laptop. Auditory stimuli were conveyed through headphones.

Each trial began with the auditory presentation of a number word. Immediately after stimulus off-set, an Arabic number appeared on the screen until participants responded, with a maximum duration of 4 s. Participants were instructed to press the green button ("l" key on the right side of the QWERTZ keyboard) when auditory and visual numbers matched and to press the red button ("a" key on the left) in case of a mismatch. The intertrial interval was 400 ms.

Numerical stimuli consisted of the digits 1–9, excluding 7 because it is the only disyllabic number word in both German and English. Visual stimuli were presented in black on a white background in Arial font with a proportional height of 0.3 compared with overall screen size. Auditory stimuli were recorded by a female native speaker in each language. All numbers were trimmed to remove excess time before and after the spoken numbers. The duration was matched by slightly adapting the speech rate to a mean level in each language. We aimed to present items at a natural speech rate while min-

---

[2] The marked loss of children in the English sample was caused by one school that did not participate in Grade 3. Results for Grade 2 were highly similar when including only children who had data from both time points.

imizing differences in duration between the two languages due to the fact that German number words contain the connector *und* (English *and*) between decades and units (e.g., *vierundzwanzig*–English *four-and-twenty*). After matching, the duration per number word was 1.4 s in German and 1.2 s in English. Altogether, 16 different auditory targets (no teens, no decade numbers, and no ties) were presented (for an overview of all items, see Supplementary Table S1 in the online supplementary material). Each verbal number word (e.g., *twenty-four*) was presented four times. To avoid a bias toward "no" responses, on 50% of the trials (*n* = 32), verbal number words were followed by the matching Arabic number. The remaining 32 trials comprised 4 possible nonmatching distractors, with each distractor type occurring 8 times: (a) an inverted distractor U+D+ (e.g., *twenty-four → 42*); (b) the decade matched with the target, whereas a different digit appeared in the unit position D+U− (e.g., *twenty-four → 28*); (c) the unit digit of the target appeared at the decade position, whereas the unit position had no overlap U+D− (e.g., *twenty-four → 48*); and (d) a nonrelated distractor D−U− (e.g., *twenty-four → 36*).

Participants received oral instructions from the experimenter. Six practice trials with feedback were included at the beginning of the task. The 64 two-digit items were combined with 24 teen numbers and 32 three-digit numbers, which were not the focus of the current study because we used different distractor items. Thus, altogether, participants were exposed to 120 items (60 matching and 60 nonmatching). The order of experimental trials was pseudorandomized with the restriction that identical number words were never presented consecutively and no more than three trials with the same expected response appeared in immediate succession.

In Grade 3, children were presented with the same cross-modal matching task with slight modifications. We added a fifth distractor type D−U+ in which the unit matched the target in correct position, whereas the decade was replaced (e.g., *twenty-four → 84*). In addition, the number of nonrelated distractors was doubled from 8 to 16. To balance the number of "yes" and "no" responses, we adjusted the number of matching targets. Therefore, children mastered 96 two-digit numbers (including 48 matches and 48 mismatches) combined with 24 teens and 48 three-digit numbers, amounting to a total number of 168 items.

Adults received the same task version as children during the Grade 2 assessment. Split-half reliability based on RTs for correct responses for two-digit items was .77 in Grade 2, .88 in Grade 3, and .89 for adults.

Arithmetic

In Grade 2, children completed selected items of the Numerical Operations subtest from the Wechsler Individual Achievement Test – Second Edition (WIAT-II UK; Wechsler, 2005). Items were adapted for group use and assimilated to language-dependent notation of arithmetic operations. First, children needed to master 6 items that involved identifying and writing Arabic digits to dictation and counting dots. Afterward, they worked on 17 standard arithmetic calculations (addition, subtraction, multiplication, and division with one- to three-digit numbers) with increasing difficulty for 15 min. Scores represent the number of correct items. Cronbach's α for this test was .76 for the current sample.

In Grade 3, children received a slightly adapted version of the Numerical Operations subtest. The first 6 items were dropped because we expected ceiling effects. The task was extended with basic arithmetic operations with three-digit and decimal numbers. Within 15 min, children worked through a total of 26 items. Cronbach's α was .84 for the current sample.

Adults were individually administered the subtests on simple additions, subtractions, and multiplications from a diagnostic inventory of arithmetic skills for elementary school age (Grube, Weberschock, Blum, & Hasselhorn, 2010). They were asked to write down their responses to the list of 110 items per calculation type. Total time to complete each subtest was recorded. Response accuracy was high (*M* = 96.97%, *SD* = 2.60). RTs were *z*-transformed (with lower *z*-scores representing poorer task performance) separately for each task and averaged. Cronbach's α for the three arithmetic operations was .88 for the adult sample.

Domain-general predictors (children only)

Nonverbal reasoning[3]. A modified version of Raven's Standard Progressive Matrices (Raven, Raven, & Court, 1998) was administered in class. The task was presented in an A5 booklet containing 36 items. The first 2 items (A1 and A2) were used for practice. The remaining items included A3 to A12, B1 to B12, and C1 to C12. The task was adapted for group administration by including the answer options on the answer sheet. One point was given for each correct response (maximum score = 34). Cronbach's α was .78 for the current sample.

Working memory[4]. The Digit Recall (forward and backward) and Block Recall subtests of the Working Memory Test Battery for Children (WMTB-C; Pickering & Gathercole, 2001) were given as measures of verbal and visual–spatial working memory. Because the Block Recall task included only a forward condition, we added a backward condition. Children were asked to tap an increasingly long sequence of blocks in reverse order compared with the experimenter. Scores of each subtest reflect the number of correctly repeated trials. *z*-Scores were computed for each subtest separately for both groups and were averaged across conditions. Reliability was checked by correlating the averaged *z*-scores for both forward tasks and both backward tasks. The correlation was $r = .44$ ($p < .001$) in both languages.

Results

The data supporting the findings of this study were collected as part of a larger longitudinal project. The whole dataset of this longitudinal project also including the data of the present study is available on ReShare, the UK Data Service's online data repository, and can be accessed at https://reshare.uk-dataservice.ac.uk/854398/.

Number-matching accuracy

No informative effects were expected for response accuracy because cross-modal matching of two-digit numbers was quite easy, even in Grade 2, with accuracy rates around 90% in all but the inverted distractor conditions, where they were around 80% (see Supplementary Table S2 for details). For the sake of completeness, results of the error analyses are reported in the online supplementary material.

Number-matching RTs

Data preprocessing

Only RTs of correct responses were analyzed. Data loss due to errors of omission was observed only in children and did not exceed 2.6% in any of the conditions. Correct responses below 200 ms were excluded because they were assumed to be anticipatory. Note that numbers were negligible (altogether, 7 items in German-speaking and 43 in English-speaking children in Grade 2 and 2 items in German-speaking and 28 in English-speaking children in Grade 3). For each group and condition, RTs more than 2.5 standard deviations above the group average were considered outliers and were excluded. This removal procedure did not affect more than 3.22% of all responses in any condition.

Exclusion of participants

This task was developed to reveal individual differences in visual–verbal transcoding of two-digit numbers in participants who had already mastered the basic steps of integrating tens and units. The overall high response accuracy rates indicate that by Grade 2 the majority of children had mastered these basic steps. Still, to guarantee validity and reliability of the RT analysis, we decided to exclude children who, after data trimming and outlier correction, had overall less than 80% of RTs or less than half of RTs in any of the distractor conditions left. In Grade 2, this exclusion procedure affected 29 German-speaking and 66 English-speaking children. Data from 119 German-speaking children (51

---

[3] This study was part of a larger longitudinal study, and these tasks were administered in Grade 1.
[4] This study was part of a larger longitudinal study, and these tasks were administered in Grade 1.

female) with a mean age of 8;2 (*SD* = 0;4) and 179 English-speaking children (79 female) with a mean age of 7;3 (*SD* = 0;4) were further analyzed. From the second assessment 1 year later in Grade 3, data from 13 German-speaking and 26 English-speaking children needed to be removed. Therefore, data from 132 German-speaking children (57 female) with a mean age of 9;1 (*SD* = 0;4) and 164 English-speaking children (76 female) with a mean age of 8;3 (*SD* = 0;4) could be analyzed. Based on the same criteria, we also needed to exclude RTs of 1 German-speaking and 3 English-speaking adults, who were probably not sufficiently attentive during task performance. Thus, mean RTs were available for 37 German-speaking adults (26 female) with a mean age of 23;5 (*SD* = 2;8) and 39 English-speaking adults (28 female) with a mean age of 19;7 (*SD* = 2;1).

Analysis of distractor conditions

Raw mean RTs are presented in Supplementary Table S3. Because we were mostly interested in the impact of distractor type on RTs, we directly compared the correct no-responses to distractors with different degrees of overlap with the target (U+D+, D+U−, U+D−, as well as D−U+ in Grade 3 only). The nonrelated distractor should be the fastest to reject in all accounts and was regarded as baseline. To control for overall RT differences between participants, raw RTs were *z*-transformed in relation to the nonrelated distractor (D−U−). To exemplify, for U+D+ this *z*-score was calculated by subtracting the mean RT of the nonrelated D−U− distractor from the mean RT of the inverted U+D+ distractor and dividing it by the standard deviation of the D−U− distractor (see Fazio, 1990). Thus, the resulting *z*-scores allowed us to investigate the cost involved in correctly rejecting different distractors regardless of differences in general RTs. This procedure also helped to rule out baseline differences in RTs due to the age difference between the two language groups for children.

Analysis of standardized RTs

Grade 2

Fig. 1 presents mean *z*-scores for each distractor condition for German- and English-speaking children in Grade 2.

z-Standardized mean RTs (*z*RTs) were subjected to a 2 × 3 repeated-measures analysis of variance (ANOVA) with language (German or English) as a between-participants factor and distractor type (U+D+, D+U−, U+D−) as a within-participants factor. Because the standardization was computed separately for the two language groups, no main effect of language was evident, $F(1, 296) < 1$. The Greenhouse–Geisser corrected ANOVA[5] showed a significant effect of distractor type, $F(2, 543) = 48.79$, $p < .001$, $\eta_p^2 = .14$, which was modulated by an interaction with language, $F(2, 543) = 21.37$, $p < .001$, $\eta_p^2 = .07$. Pairwise comparisons of distractor conditions using Bonferroni correction for multiple comparisons (the Bonferroni-corrected α level was .008) revealed that, as expected, German-speaking children had the highest *z*-score in the inversion distractor condition U+D+ (all *p*s < .001). When just one digit overlapped with the verbally presented stimulus, the unit in the first position of the visual number (U+D−) distracted German-speaking children significantly more than the decade (D+U−) ($p = .001$). One-sample *t* tests confirmed that all three distractors differed significantly from 0, which corresponds to the *z*-score of the nonrelated distractor condition, $t(118)$ scores between 2.54 and 8.47, *p*s between <.001 and .012, *d*s between 0.23 and 0.78. To sum up, German-speaking children showed the following pattern: U+D+ > U+D− > D+U− > D−U−.

For English-speaking children, a different pattern appeared. Again, the difference between inverted distractors (U+D+) and overlapping unit distractors (U+D−) was significant ($p = .002$), and the difference between inverted distractors (U+D+) and decade-related distractors (D+U−) was marginally significant ($p = .049$) (the Bonferroni-corrected α level was .008). Interestingly, the two distractors overlapping in only one digit (U+D− and D+U−) did not differ significantly in the English Grade 2 sample ($p = .140$). Again, one-sample *t* tests confirmed that mean *z*RTs of all distractor conditions were significantly different from 0, representing the unrelated condition: $t(178)$ scores between 5.85 and

---

[5] In case Mauchly's test indicated a violation of the sphericity assumption, ANOVAs were Greenhouse–Geisser corrected.
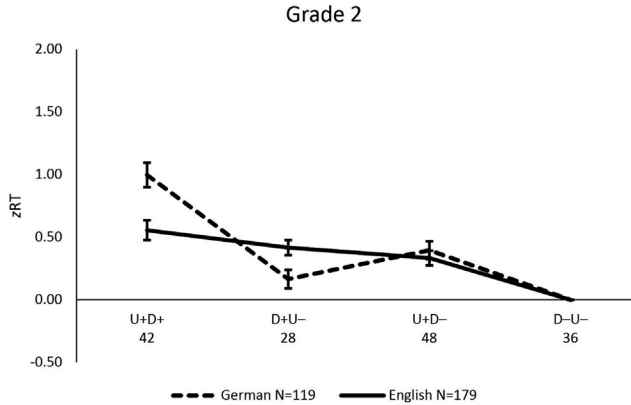
Fig. 1. *z*-Standardized mean RTs (zRTs) for four different distractor types (U+D+, D+U−, U+D−, and D−U−) in German-speaking (dashed line) and English-speaking (solid line) second-graders. The presented two-digit numbers below the distractor labels exemplify the distractors for the verbally presented number *twenty-four*. Error bars represent standard errors.

8.35, all *p*s < .001, *d*s between 0.44 and 0.62. The resulting pattern for English-speaking children was U +D+ > U+D− but U+D+ ≈ D+U− and D+U− ≈ U+D− (all overlapping distractors > D−U−).

Importantly, German-speaking children showed significantly stronger interference of the inverted distractor (U+D+) than English-speaking children (*p* = .001) (the Bonferroni-corrected α level was .017), whereas English-speaking children showed stronger interference of the decade-consistent distractor D+U− (*p* = .009). For the unit-related distractor condition U+D−, no language difference was observed (*p* = .503).

Grade 3

The repetition of the number-matching task 1 year later in Grade 3 confirmed and extended the Grade 2 results (see Fig. 2).

A 2 × 4 repeated-measures ANOVA with language (German or English) as a between-participant factor and distractor type (U+D+, D+U−, U+D−, D−U+) as a within-participants factor revealed main effects of language, $F(1, 294) = 10.80$, $p = .001$ $\eta_p^2 = .04$, and distractor type, $F(3, 789) = 145.56$, $p > .001$, $\eta_p^2 = .33$, as well as an interaction, $F(3, 789) = 22,47$, $p = .001$, $\eta_p^2 = .07$. Pairwise comparisons (the Bonferroni-corrected α level was .004) showed that German-speaking children again experienced the strongest interference in the inverted distractor condition (all *p*s < .001). Standardized RTs for distractors including a unit were again significantly higher than those for decade-consistent distractors D +U− ($p ≤ .001$). Interestingly, this was also the case for the newly added D−U+ distractor, whereas no difference was observed between U+D− and D−U+ (*p* = .772), indicating that the position of the overlapping unit was irrelevant. One sample *t* tests revealed that the inversion distractor and both unit overlapping distractors differed significantly from 0, representing the nonrelated target, *t*(131) scores between 4.48 and 13.37, all *p*s < .001, *d*s between 0.39 and 1.16, whereas the decade-consistent distractor D+U− did not, *t*(131) = 1.28, *p* = .204, *d* = 0.11. In summary, the performance pattern of German-speaking children in Grade 3 was U+D+> U+D− ≈ D−U+ > D+U− ≈ D−U−.

English-speaking children also showed a pattern similar to Grade 2 with the exception that *z*RTs for inverted distractors (U+D+) were significantly larger than decade-consistent distractors D+U−, which in turn showed significantly larger *z*RTs than both unit-related distractors (U+D− and D−U+) (all *p*s < .001). Similar to the German-speaking sample, the newly added D−U+ distractor was not different from the other unit-related distractor U+D− (*p* = .282). As a matter of fact, only the U+D+ and D+U− conditions differed significantly from 0 (representing the nonrelated distractor D−U−) [U+D+: *t* (163) = 9.74, *p* < .001, *d* = 0.76; D+U−: *t*(163) = 5.64, *p* < .001, *d* = 0.44], whereas both unit-related conditions did not [U+D−: *t*(163) = 0.84, *p* = .401, *d* = 0.07; D−U+: *t*(163) = −0.43, *p* = .665, *d* = −0.03]. The overall pattern for English-speaking children in Grade 3 was U+D+ > D+U− > U+D− ≈ D−U− ≈ D−U+.
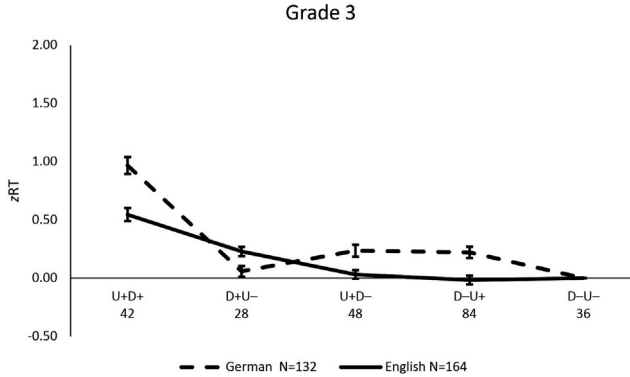
Fig. 2. *z*-Standardized mean RTs (zRTs) for five different distractor types (U+D+, D+U−, U+D−, D−U+, and D−U−) in German-speaking (dashed line) and English-speaking (solid line) third-graders. The presented two-digit numbers below the distractor labels exemplify the distractors for the verbally presented number *twenty-four*. Error bars represent standard errors.

The direct comparison of the two language groups showed that German-speaking children again had higher *z*RTs than English-speaking children for all three distractors, including the correct unit digit (all *p*s ≤ .001) (the Bonferroni-corrected α level was .013). Only in the decade-consistent condition D +U− were *z*RTs higher for English-speaking children than for German-speaking children (*p* = .005).

Adults

Fig. 3 presents the mean *z*-scores for the four distractor conditions for German- and English-speaking adults.

A 2 × 3 ANOVA again revealed a significant main effect of distractor type, $F(2, 148) = 27.55$, $p < .001$, $\eta_p^2 = .27$, and a significant interaction with language, $F(2, 148) = 8.29$, $p < .001$, $\eta_p^2 = .10$ [but no language effect due to *z*-transformation, $F(1, 74) = 0.41$, $p = .527$]. This suggests that, even among adults, RTs to different distractor types were influenced by language-specific number–word structure. Pairwise comparisons (the Bonferroni-corrected α level was .008) revealed that German-speaking adults still showed the strongest interference in response to inverted distractors (all *p*s < .001). In contrast to German-speaking children, no difference was observed between the two distractor conditions overlapping in only one digit (D+U− and U+D−, *p* = .623). One-sample *t* tests revealed that the inversion distractor U+D+ and the decade-consistent distractor D+U− differed significantly from 0, representing the nonrelated target [U+D+: $t(36) = 8.00$, $p < .001$, $d = 1.31$; D + U−: $t(36) = 4.23$, $p < .001$, $d = 0.70$], whereas the unit-related condition did not [U+D−: $t(36) = 2.50$, $p = .017$, $d = 0.41$]. Because adult sample sizes were relatively small, we wanted to rule out that the nonsignificant effect was due to lack of statistical power. We used Bayes factor (BF) analysis (Dienes, 2014), which is less dependent on sample size than standard frequentist analyses, to determine the relative strength of the null hypothesis. The BF is defined as the ratio of the likelihood of a particular hypothesis H1 to the likelihood of the null hypothesis H0 (Wagenmakers et al., 2018). Bayesian statistics were performed with JASP software (JASP Team, 2019). To interpret the BFs, we drew on standard values of BF less than 1/3 (meaning that the H0 is three times more likely than the H1) as evidence in favor of the null hypothesis and BF greater than 3 (meaning that the H1 is three times more likely than the H0) as evidence for the alternative hypothesis (Dienes, 2014; Wagenmakers et al., 2018). For German-speaking adults, a Bayesian paired-samples *t* test revealed strong evidence that *z*RTs for partly overlapping distractors (D+U− and U+D−) did not differ from each other ($BF_{10} = 0.212$). Bayesian one-sample *t* tests showed anecdotal to moderate evidence ($BF_{10} = 2.674$) in favor of larger *z*RTs in the unit-related distractor (U+D−) compared with 0, representing the nonrelated distractor. The overall pattern for German-speaking adults was U+D+ > D+U− ≈ U +D−; U+D+ / D+U− > D−U−, but U+D− ≥ D−U−.
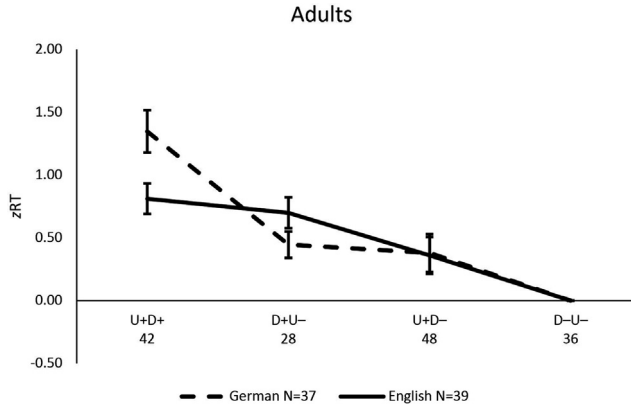
Fig. 3. z-Standardized mean RTs (zRTs) for four different distractor types (U + D+, D + U−, U + D−, and D − U−) in German-speaking (dashed line) and English-speaking (solid line) adults. The presented two-digit numbers below the distractor labels exemplify the distractors for the verbally presented number *twenty-four*. Error bars represent standard errors.

For English-speaking adults, the post hoc comparisons showed no significant differences between U+D+ and D+U− (p = .406). However, the inverted distractor (U+D+) differed significantly from the U+D− condition (p = .003). After Bonferroni correction (α level = .008), the decade-consistent distractor (D+U−) and the unit-related condition (U+D−) did not significantly differ from each other (p = .012), similar to findings in Grade 2. To check whether the three conditions were harder to reject than the unrelated D−U− condition, we ran one-sample t tests. It turned out that for English adults, only the U+D+ and D+U− conditions differed significantly from 0 (representing the nonrelated distractor D−U−), ts(38) = 6.67 and 5.71, ps < .001, ds = 1.07 and 0.91, whereas the U+D− condition did not, t(38) = 2.45, p = .019, d = 0.39.

Bayesian statistics provided substantial evidence for the null hypothesis that zRTs for inversion distractors (U+D+) and decade-consistent distractors (D+U−) were similar (BF_{10} = 0.236). The evidence that zRTs of both partly overlapping distractors (D+U− and U+D−) differed from each other can be rated as anecdotal (BF_{10} = 1.507). The same was true when we tested zRTs of the unit-related distractor U + D− against 0 (representing the nonrelated distractor) (BF_{10} = 2.368). The overall pattern for English-speaking adults was U+D+ ≈ D+U− > D−U−; U+D+ > U+D−; D+U− ≈ U+D−; U+D− ≈ D−U−. The Language × Distractor Type interaction was due to a significantly stronger inversion effect in German than in English, even among adults (p = .011) (the Bonferroni-corrected α level was .017). No language differences were found for the two other distractors (D+U−: p = .122; U + D−: p = .930).

Associations of number matching with arithmetic

To examine the association of our number-matching performance with arithmetic, two indicators of transcoding efficiency were investigated: (a) mean RTs for correct "yes" responses to matching trials (RT_match) and (b) as a direct measure of the cost of inversion, the individual zRTs of our U+D+ distractor that resulted from relating the inversion condition to the unrelated condition [inversion effect = $(M_{U+D+} - M_{D-U-}) / SD_{D-U-}$]. Our large samples of children also allowed us to investigate whether transcoding performance predicted arithmetic skills over and above domain-general predictors (age, nonverbal reasoning, and working memory). The online supplementary material presents descriptives and zero-order correlations of all variables included in the regression model for Grade 2 (Supplementary Table S4) and Grade 3 (Supplementary Table S5).

The specific association between the number-matching indicators and arithmetic was analyzed based on hierarchical regressions with age and domain-general predictors (nonverbal reasoning and working memory) entered as predictors in Step 1. As the second step, either RTs for the matching condition (Step 2A: RT_match) or the specific inversion indicator (Step 2B: inversion effect) was entered.

In Grade 2, RTs for correct matching decisions predicted arithmetic performance over and above domain-general predictors, whereas the inversion effect explained unique variance in German second-graders only (see Table 1). The language-specific effect in German highlighted the impact of number word morphology on transcoding and demonstrated its consequences for higher skills such as arithmetic performance. This explicit relationship seems to be limited to early development given that zero-order correlations between the inversion effect and arithmetic were not significant in Grade 3 (Supplementary Table S4) . However, in both languages the ability to identify a match (which contained the correct application of the inversion rule in German) turned out to be a more stable predictor of arithmetic skills (see Table 1 for Grade 3).

To consider whether the associations observed among children persisted into adulthood, simple correlations between the two transcoding indexes and arithmetic were calculated for the adult samples. In German-speaking adults, the speed to correctly detect a matching target (RT_match: $M = 534$ ms, $SD = 77$) was not significantly related to arithmetic skills ($r = -.25, p = .144$), with Bayesian statistics showing only anecdotal evidence ($BF_{10} = 0.571$). For English-speaking adults, the correlation between arithmetic and RTs for matching targets (RT_match: $M = 504$ ms, $SD = 67$) was significant ($r = -.51, p = .001$), with strong evidence in favor of the alternative hypothesis displayed by Bayesian statistics ($BF_{10} = 37.765$). The inversion effect was not significantly related to arithmetic either in German ($r = -.01, p = .976, BF_{10} = 0.205$) or in English ($r = .14, p = .388, BF_{10} = 0.286$). The BFs associated with these correlation coefficients clearly indicate that the null hypothesis of absent association between the inversion effect and arithmetic is four to five times more likely than the alternative hypothesis of an existing relationship.

Discussion

The current study investigated the relevance of linguistic and visual characteristics on transcoding of two-digit numbers in children and adults. In our number-matching paradigm, we were particularly

Table 1
Hierarchical multiple regression analyses for German- and English-speaking second-graders predicting arithmetic performance in Grade 2 and Grade 3, with the ability to reject an inverted distractor (Step 2A) and the ability to identify a match (Step 2B).

	Grade 2						Grade 3					
	German-speaking children			English-speaking children			German-speaking children			English-speaking children		
	t	β	sr^2	t	β	sr^2	t	β	sr^2	t	β	sr^2
Step 1												
Age	2.26*	.19	.04	2.04*	.15	.02	2.02*	.15	.02	1.51	.10	.01
Nonverbal reasoning	2.87**	.26	.06	2.18*	.16	.02	4.49***	.36	.11	3.92***	.28	.06
Working memory	2.00*	.18	.03	3.70***	.27	.06	3.37**	.27	.06	5.69***	.39	.14
R^2	.19***			.18***			.32***			.39***		
Step 2A												
RT_match	−3.67***	−.31	.09	−2.84**	−.20	.04	−3.33**	−.24	.05	−3.58***	−.23	.05
Total R^2	.28***			.22***			.38***			.39***		
ΔR^2	.09***			.04**			.06**			.05***		
Step 2B												
Inversion effect[a]	−2.09*	−.18	.03	−1.54	−.11	.01	−0.92	−.07	<.01	−0.34	−.02	<.01
Total R^2	.22***			.19***			.33***			.34***		
ΔR^2	.03*			.01			<.01			<.01		
n	118			178			131			160		

Note. The dependent variable is the Numerical Operations subtest of the WIAT-II UK. The beta coefficients from domain-general predictors are reported only for Step 1 because there was just a minimal change in beta coefficients from Step 1 to Step 2. sr^2, squared semipartial correlations.
* $p < .05$.
** $p < .01$.
*** $p < .001$.
[a] Inversion effect = $(M_{U+D+} - M_{D-U-}) / SD_{D-U-}$.

interested in (a) the linguistic–morphological effect of the inversion property in German, but not English, spoken number words and (b) the effect of visual overlap between the Arabic digit string representing the spoken number word and the presented Arabic digit string of distractor items. By comparing specifically selected distractor types, we aimed at differentiating among three different explanatory accounts (lexicalization, visual, and linguistic–morphological accounts). An important feature of the current paradigm is that it allows analysis of RTs and thus is not influenced by ceiling effects. As a matter of fact, we specifically aimed at investigating verbal–visual transcoding at times when the integration of tens and units into two-digit numbers has already been acquired. The systematic RT differences between distractor conditions are not in line with the assumption of the ADAPT model (Barrouillet et al., 2004) that two-digit numbers are lexicalized early in development. Differences between distractor conditions were evident in German, a language with number–word inversion, as well as in the noninverted English number language across all age levels, pointing toward procedural transcoding rather than lexical retrieval.

Within each language, the performance patterns across distractor conditions were surprisingly similar among children in both Grades 2 and 3 and even among adults. The patterns unveiled persistent difficulties in rejecting the U+D+ distractor in German compared with English, which can be best explained by the German number–word structure. The unit number appears first in the auditory number word (e.g., *vier-und-zwanzig* literally means *four-and-twenty*), which mirrors the order of digits in the common left-to-right reading direction. These language effects indicate that the inversion property places specific processing demands during simple transcoding procedures even for double-digit numbers. This evidence corroborates earlier findings indicating that French-speaking children responded faster to two-digit numbers than German-speaking children who need to deal with inversion (Poncin et al., 2019). It extends earlier reports on differences in transcoding performance between children speaking languages with different number–word structures (Imbo et al., 2014; Moeller, Zuber et al., 2015; Poncin et al., 2019; Van Rinsveld & Schiltz, 2016) because it is the first systematic cross-linguistic comparison of the highly similar German and English languages that revealed an effect of number–word structure even in highly experienced adults. Our findings indicate that the inversion property adds specific processing demands during simple transcoding procedures even for double-digit numbers. Whereas earlier studies on inversion mostly reported the unsurprising finding that when young children read or write multidigit numbers, inverted languages provoke more inversion errors than noninverted languages, we took a critical further step by testing competing accounts of the mechanisms underlying audio–visual number matching.

The performance pattern of German-speaking children in Grades 2 and 3 fits perfectly with the assumption of a linguistic–morphological account; the inversion condition (U+D+) constituted a particular challenge, with significantly higher RTs than in all other conditions. Moreover, German-speaking children showed higher RTs in the U+D− condition than in the D+U− condition. Thus, both distractors containing the target unit (U+D+ and U+D−) induced higher RTs than the decade-related distractor (D+U−). Interestingly, decade-related distractors could be rejected just as quickly as unrelated distractors in the Grade 3 assessment, suggesting that children were able to reject as soon as they realized that the perceived unit number was not contained in the Arabic double digit.

Note that our findings do not provide much support for the idea that German individuals may adapt their direction of processing Arabic numbers to a right-to-left fashion in order to process the unit first. None of the distractor conditions given in Grade 2 included the correct unit on the rightmost position, so based on a right-to-left processing direction, they all could be rejected right away without major differences between conditions. More evidence against clear right-to-left processing comes from the fact that we did not observe differences between U+D− and D−U+ conditions in our Grade 3 assessment. It seems that the presence of the matching unit number delayed rejection (compared with English-speaking peers) regardless of its exact position.

Results for German-speaking adults were somewhat less clear than those for children given that the difference between the D+U− and U+D− conditions was not significant, resembling the pattern for English-speaking adults. Note that Poncin et al. (2019) also used an audio–visual matching paradigm with German-speaking 10-year-olds and adults. In their task, participants needed to select the matching number out of four Arabic options. Interestingly, adults responded faster when the units

of the response alternatives were presented with a 500-ms delay than when the units were presented first, whereas 10-year-olds did not (yet) seem to profit from a decade-first presentation. Poncin et al. suggested that transcoding may evolve to a sequential ten-unit order over time, but this development might be slower in a language with inversion. Such a developmental perspective would be in line with the performance patterns observed here.

In general, English-speaking participants also showed interference for inverted distractors. However, the zRTs were consistently smaller than in the matched German-speaking samples. Interestingly, English-speaking second-graders and adults not only showed no difference between unit-related and unrelated distractors, but they also did not need more time to reject the inverted distractor compared with the decade-consistent distractor D+U−. Note that a purely visual account would predict differences between those conditions, as well as between the two unit-related conditions, due to differences in visual overlap. On the other hand, the D+U− and U+D− conditions are relatively similar in their amount of visual overlap (both have one digit in common with the target) but clearly differed in RTs across all age groups. Only in the English Grade 3 assessment did we observe significantly higher RTs for the inverted distractor compared with the visually less overlapping decade-consistent distractor, indicating that visual features are also at play in cross-modal number matching.

Importantly, the overall RT patterns for English-speaking children were profoundly different from those observed in German. In Grade 2, inverted items (U+D+) took longer to reject than U+D− items overlapping only in unit identity. The D+U− condition, overlapping in decade identity as well as position, also induced high interference, which was actually not different from either the inverted condition (U+D+) or the unit-consistent position (U+D−). One year later, in Grade 3, both distractors including the decade (U+D+ and D+U−) caused significantly higher RTs than both distractors including the unit (U+D− and D−U+), which no longer differed from the unrelated distractor (D−U−). Thus, at this age any visual overlap in the unit position no longer affected children's decision to reject a nonmatching item. This can be interpreted as further evidence for a linguistic–morphological account: Because English number words start with the decade, all items that do not include the decade number can be disregarded quickly regardless of any unit overlap.

Our findings demonstrate the crucial relevance of a cross-linguistic approach directly comparing languages with deviant number–word structure. Solely based on the performance pattern in English, a clear differentiation between a visual account and a linguistic–morphological account would be difficult because they make similar predictions for languages with consistent order of the constituents of verbal and visual Arabic numbers. Importantly, the differences in RT patterns between age groups indicate that the development of transcoding continues even after correct application of transcoding rules.

The audio–visual number-matching experiment specifically addressed mechanisms involved in transcoding. The task format (presentation of spoken number words followed by Arabic digit strings) encouraged participants to successively analyze and convert spoken number words into Arabic digits as proposed by the ADAPT model (Barrouillet et al., 2004). The model was originally designed for number writing. Our task format did not involve explicit number production but rather involved verification of written Arabic numbers, which also requires transcoding from verbal numbers into Arabic numbers. The current results show that the ADAPT model needs some adaptation. As pointed out by others (Imbo et al., 2014; Zuber et al., 2009), it should be adjusted to explain linguistic characteristics affecting transcoding procedures. Languages with inversion property require an additional procedural step that is currently lacking from the model. The assumption of lexicalization should be revisited given that we found a long-lasting impact of language on cross-modal matching. Even though the linguistic–morphological account was the most plausible approach to explain differences between English- and German-speaking participants, we cannot entirely rule out that lexicalization and visual characteristics might contribute to performance on the number-matching task. We do not see the three accounts discussed here as mutually exclusive.

Based on the verbal–visual matching paradigm, we also aimed to contribute to the ongoing debate on whether or not transcoding performance is associated with arithmetic performance. Although such an association has repeatedly been shown for the very first steps of understanding multidigit numbers (Göbel, Moeller, et al., 2014; Göbel, Watson, et al., 2014; Imbo et al., 2014; Moeller et al., 2011; Purpura, Baroody, & Lonigan, 2013), findings are inconsistent for transcoding efficiency beyond early

development, as assessed in terms of RTs (Lin & Göbel, 2019; Lyons et al., 2014; Sasanguie et al., 2017; Sasanguie & Reynvoet, 2014). Furthermore, no other study so far has investigated this association cross-linguistically, directly comparing two languages with inverted versus consistent number word structure.

For children in both languages, we found significant associations of arithmetic performance with the efficiency to identify an Arabic double-digit number matching a previously presented number word even after controlling for domain-general predictors. Thus, at the end of Grade 2, as well as at the end of Grade 3, the efficiency of matching verbal and visual Arabic representations of multidigit numbers is a relevant foundation of children's arithmetic performance regardless of linguistic number–word structure. Even though our adult samples were much smaller because we mostly wanted to investigate long-term impacts of inversion on transcoding itself, we still found a significant association between transcoding efficiency and arithmetic for English, although not for German. For two-digit numbers, identifying a match in German involves the imperative application of inversion rule. In our arithmetic task, however, participants performed simple arithmetic exercises containing many single-digit operations. This might explain the weak but nonsignificant relationship. Future studies should investigate this association in larger samples and with more sensitive measures of arithmetic performance.

We also explored whether the efficiency with which participants processed decade–unit inversion was related to their arithmetic performance by relating the specific cost to reject an inverted number with arithmetic performance. Interestingly, even after controlling for domain-general predictors, this association was significant for German-speaking, but not English-speaking, second-graders. This provides tentative evidence that efficient processing of decade–unit inversion is more relevant to arithmetic in a language that systematically implements the inversion property. This result is consistent with the findings by Göbel, Moeller, et al. (2014), who reported more processing time for additions with carry in German second-graders than for Italian second-graders.

One year later in Grade 3, as well as in our relatively small adult samples, we did not observe significant correlations between the inversion effect and arithmetic in either German or English. Thus, inversion clearly influences number processing but might have rather temporary and specific effects on arithmetic performance.

Conclusion

Our findings support the idea that the process of transcoding two-digit numbers is driven by language-specific number–word structure. German-speaking participants who need to deal with the reversed order of units and tens showed higher interference when rejecting an Arabic number containing the unit. The opposite pattern was found for English-speaking participants who showed higher interference when the decade was included. These findings provide evidence against early lexicalization of two-digit number words as proposed by the ADAPT model (Barrouillet et al., 2004) and cannot be explained by a visual account. This is the first evidence that theoretical models of transcoding need to account for language-specific transcoding rules such as inversion (Zuber et al., 2009) even in skilled transcoders. Moreover, proficient switching between number words and Arabic numbers indicated by faster identification of a verbal–visual match was related to better arithmetic skills. These findings illustrate the relevance of multidigit decoding efficiency for arithmetic performance, especially during development.

Acknowledgments

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jecp.2020.104970.

References

Barrouillet, P., Camos, V., Perruchet, P., & Seron, X. (2004). ADAPT: A developmental, asemantic, and procedural model for transcoding from verbal to Arabic numerals. *Psychological Review, 111*, 368–394.

Chrisomalis, S. (2010). *Numerical notation: A comparative history*. Cambridge, UK: Cambridge University Press.

Cohen, D. J. (2009). Integers do not automatically activate their quantity representation. *Psychonomic Bulletin & Review, 16*, 332–336.

Cohen, D. J., Warren, E., & Blanc-Goldhammer, D. (2013). Cross-format physical similarity effects and their implications for the numerical cognition architecture. *Cognitive Psychology, 66*, 355–379.

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition, 44*, 1–42.

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology, 5*. https://doi.org/10.3389/fpsyg.2014.00781.

Dotan, D., & Friedmann, N. (2018). A cognitive model for multidigit number reading: Inferences from individuals with selective impairments. *Cortex, 101*, 249–281.

Dowker, A., Bala, S., & Lloyd, D. (2008). Linguistic influences on mathematical development: How important is the transparency of the counting system? *Philosophical Psychology, 21*, 523–538.

Fazio, R. H. (1990). A practical guide to the use of response latency in social psychological research. In C. Hendrick & M. S. Clark (Eds.), *Research methods in personality and social research* (pp. 74–97). Newbury Park, CA: Sage.

Göbel, S. M., Moeller, K., Pixner, S., Kaufmann, L., & Nuerk, H. C. (2014). Language affects symbolic arithmetic in children: The case of number word inversion. *Journal of Experimental Child Psychology, 119*, 17–25.

Göbel, S. M., Watson, S. E., Lervåg, A., & Hulme, C. (2014). Children's arithmetic development: It is number knowledge, not the approximate number sense, that counts. *Psychological Science, 25*, 789–798.

Grube, D., Weberschock, U., Blum, M., & Hasselhorn, M. (2010). *Diagnostisches Inventar zu Rechenfertigkeiten im Grundschulalter (DIRG) [Diagnostic Inventory of Arithmetic Skills at Elementary School Age]*. Göttingen, Germany: Hogrefe.

Huber, S., Mann, A., Nuerk, H. C., & Moeller, K. (2014). Cognitive control in number magnitude processing: Evidence from eye-tracking. *Psychological Research, 78*, 539–548.

Imbo, I., Vanden Bulcke, C., De Brauwer, J., & Fias, W. (2014). Sixty-four or four-and-sixty? The influence of language and working memory on children's number transcoding. *Frontiers in Psychology, 5*. https://doi.org/10.3389/fpsyg.2014.00313.

JASP Team. (2019). JASP (Version 0.11.1) [computer software]. Retrieved from https://jasp-stats.org.

Lin, C.-Y., & Göbel, S. M. (2019). Arabic digits and spoken number words: Timing modulates the cross-modal numerical distance effect. *Quarterly Journal of Experimental Psychology, 72*, 2632–2646.

Lyons, I. M., Price, G. R., Vaessen, A., Blomert, L., & Ansari, D. (2014). Numerical predictors of arithmetic success in Grades 1–6. *Developmental Science, 17*, 714–726.

McCloskey, M. (1992). Cognitive mechanisms in numerical processing: Evidence from acquired dyscalculia. *Cognition, 44*, 107–157.

Moeller, K., Huber, S., Nuerk, H.-C., & Willmes, K. (2011). Two-digit number processing: Holistic, decomposed or hybrid? A computational modelling approach. *Psychological Research, 75*, 290–306.

Moeller, K., Pixner, S., Zuber, J., Kaufmann, L., & Nuerk, H.-C. (2011). Early place-value understanding as a precursor for later arithmetic performance—A longitudinal study on numerical development. *Research in Developmental Disabilities, 32*, 1837–1851.

Moeller, K., Shaki, S., Göbel, S. M., & Nuerk, H.-C. (2015). Language influences number processing—A quadrilingual study. *Cognition, 136*, 150–155.

Moeller, K., Zuber, J., Olsen, N., Nuerk, H.-C., & Willmes, K. (2015). Intransparent German number words complicate transcoding—A translingual comparison with Japanese. *Frontiers in Psychology, 6*. https://doi.org/10.3389/fpsyg.2015.00740.

Moura, R., Wood, G., Pinheiro-Chagas, P., Lonnemann, J., Krinzinger, H., Willmes, K., & Haase, V. G. (2013). Transcoding abilities in typical and atypical mathematics achievers: The role of working memory and procedural and lexical competencies. *Journal of Experimental Child Psychology, 116*, 707–727.

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

Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods, 162*, 8–13.

Pickering, S. J., & Gathercole, S. E. (2001). *Working memory test battery for children (WMTB-C)*. London: Psychological Corporation.

Pixner, S., Zuber, J., Heřmanová, V., Kaufmann, L., Nuerk, H.-C., & Moeller, K. (2011). One language, two number-word systems and many problems: Numerical cognition in the Czech language. *Research in Developmental Disabilities, 32*, 2683–2689.

Poncin, A., Van Rinsveld, A., & Schiltz, C. (2019). Units-first or tens-first: Does language matter when processing visually presented two-digit numbers? *Quarterly Journal of Experimental Psychology, 73*, 726–738.

Power, R. J. D., & Dal Martello, M. F. (1990). The dictation of Italian numerals. *Language and Cognitive Processes, 5*, 237–254.

Power, R. J. D., & Dal Martello, M. F. (1997). From 834 to eighty thirty four: The reading of Arabic numerals by seven-year-old children. *Mathematical Cognition, 3*, 63–85.

Purpura, D. J., Baroody, A. J., & Lonigan, C. J. (2013). The transition from informal to formal mathematical knowledge: Mediation by numeral knowledge. *Journal of Educational Psychology, 105*, 453–464.

Raven, J., Raven, J. C., & Court, J. H. (1998). *Raven's standard progressive matrices and vocabulary scales*. Oxford, UK: Oxford Psychologists Press.

Reeve, R. A., Reynolds, F., Paul, J., & Butterworth, B. L. (2018). Culture-independent prerequisites for early arithmetic. *Psychological Science, 29*, 1383–1392.

Sasanguie, D., Lyons, I. M., De Smedt, B., & Reynvoet, B. (2017). Unpacking symbolic number comparison and its relation with arithmetic in adults. *Cognition, 165*, 26–38.

Sasanguie, D., & Reynvoet, B. (2014). Adults' arithmetic builds on fast and automatic processing of Arabic digits: Evidence from an audiovisual matching paradigm. *PLoS One, 9*(2) e87739.

Simmons, F. R., Willis, C., & Adams, A. M. (2012). Different components of working memory have different relationships with different mathematical skills. *Journal of Experimental Child Psychology, 111*, 139–155.

van der Ven, S. H. G., Klaiber, J. D., & van der Maas, H. L. J. (2017). Four and twenty blackbirds: How transcoding ability mediates the relationship between visuospatial working memory and math in a language with inversion. *Educational Psychology, 37,* 487–505.

Van Rinsveld, A., & Schiltz, C. (2016). Sixty-twelve = Seventy-two? A cross-linguistic comparison of children's number transcoding. *British Journal of Developmental Psychology, 34,* 461–468.

van Viersen, S., Slot, E. M., Kroesbergen, E. H., van't Noordende, J. E., & Leseman, P. P. M. (2013). The added value of eye-tracking in diagnosing dyscalculia: A case study. *Frontiers in Psychology, 4.* https://doi.org/10.3389/fpsyg.2013.00679.

Wagenmakers, E. J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., ... Morey, R. D. (2018). Bayesian inference for psychology: I. Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review, 25,* 35–57.

Wechsler, D. (2005). *Wechsler Individual achievement test-second UK edition (WIAT-IIUK).* London: Harcourt Assessment.

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