Article:

Bahnmueller, Julia, Maier, Carolin A, Göbel, Silke M. orcid.org/0000-0001-8845-6026 et al. (1 more author) (2018) Direct evidence for linguistic influences in two-digit number processing. Journal of Experimental Psychology: Learning, Memory, and Cognition. ISSN 1939-1285

https://doi.org/10.1037/xlm0000642

To appear in: Journal of Experimental Psychology: Learning, Memory and Cognition

Direct evidence for linguistic influences in two-digit number processing

Julia Bahnmueller

Leibniz-Institut für Wissensmedien, Tuebingen, Germany and University of Tuebingen, Germany, and LEAD Graduate School and Research Network, University of Tuebingen, Germany

Carolin A. Maier

Leibniz-Institut für Wissensmedien, Tuebingen, Germany, International Max Planck Research School on Neuroscience of Communication: Function, Structure and Plasticity, Leipzig, Germany, and Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

Silke M. Göbel

University of York, United Kingdom

Korbinian Moeller

Leibniz-Institut für Wissensmedien, Tuebingen, Germany, and University of Tuebingen, Germany and LEAD Graduate School and Research Network, University of Tuebingen, Germany

Author Note

Julia Bahnmueller, Neuro-Cognitive Plasticity Laboratory, Leibniz-Institut für Wissensmedien, Tuebingen, Germany; Carolin A. Maier, Neuro-Cognitive Plasticity Laboratory, Leibniz-Institut für Wissensmedien, Tuebingen, Germany; Silke M. Göbel, Department of Psychology, University of York, United Kingdom; Korbinian Moeller, Neuro-Cognitive Plasticity Laboratory, Leibniz-Institut für Wissensmedien, Tuebingen, Germany.

Julia Bahnmueller was supported by the Leibniz-Competition Fund (SAW-2014-IWM-4). We would like to thank Charlotte Ashton, Lia Heubner, and Marie-Lene Schlenker for help with data collection.

Correspondence concerning this article should be addressed to Julia Bahnmueller, Leibniz-Institut für Wissensmedien, Schleichstraße 6, 72076 Tuebingen, Germany. Email: j.bahnmueller@iwm-tuebingen.de

Abstract

Language-specific differences in number words influence number processing even in non-verbal numerical tasks. For instance, the unit-decade compatibility effect in two-digit number magnitude comparison (compatible number pairs [42_57: 4<5 and 2<7] are responded to faster than incompatible pairs [47_62: 4<6 but 7>2]) was shown to be influenced by the inversion of number words (e.g., in German the number word for 42 is *zweiundvierzig* [literally: *two-and-forty*]). In two studies, we used articulatory suppression to investigate whether previously observed cross-linguistic differences in two-digit number processing are indeed driven by differences in number word formation. In a two-digit number comparison task, German- and English-speaking participants had to identify the larger of two numbers presented in Arabic digits. In Study 1, participants performed the same task twice, with and without articulatory suppression. In Study 2, the percentage of within-decade filler items (36_39) was manipulated additionally. As expected, in both studies between-group differences in the compatibility effect disappeared under articulatory suppression irrespective of the percentage of fillers included. Furthermore, paralleling results of previous studies including 33% or less filler items, we found that the compatibility effect was larger in German compared to English speakers in the 20% filler condition. However, this pattern was reversed in the 50% filler condition in both studies. Thus, results provide first direct evidence for influences of verbal number word formation on symbolic number processing. Moreover, these new findings suggest that linguistic influences and those of cognitive control processes associated with characteristics of the stimulus set interact in symbolic number processing.

Word count: 250

Keywords

Compatibility effect, linguistic influences, articulatory suppression, filler manipulation, cognitive control

Introduction

International studies evaluating scholastic abilities (e.g., reading, writing and mathematics) have consistently reported large cross-cultural differences in mathematical performance (e.g., OECD, 2014). In addition to differences in schooling and cultural valuation it has been argued that the language of mathematics learning and instruction matters as well. For example, the degree of correspondence between the Arabic number system and a language's number word system may partly account for the observed cross-cultural differences (e.g., Moeller, Shaki, Göbel, & Nuerk, 2015; see Okamoto, 2015 for a review). In this context, previous research highlighted the relative simplicity of Chinese number words, also called the "Chinese number advantage" (e.g., Miller, Kelly, & Zhou, 2005; Miura & Okamoto, 2003). Importantly, both power and order are expressed transparently in both the number and the number word system (see Miller et al., 2005; Ngan Ng & Rao, 2010 for more detailed descriptions). However, for other languages, (e.g., German, Dutch, Maltese, etc. see Comrie, 2005) the consistency between number words and Arabic digits is not as high as one would expect when looking at the highly systematic structure of the Arabic number system. Especially the counterintuitive inversion of tens and units in several number word systems (e.g., 25 → "fünfundzwanzig" – literally five and twenty in German, but also in Dutch, Arabic, Maltese, etc.) not only constitutes an obstacle during the acquisition of multi-digit number knowledge (e.g., Göbel, Moeller, Pixner, Kaufmann, & Nuerk , 2014; Helmreich et al., 2011; Imbo, Vanden Buckle, De Bauwer, & Fias, 2014; Xenidou-Dervou, Gilmore, van der Schoot, & van Lieshout, 2015) but also influences number processing for numerically skilled adults (e.g., Brysbaert, Fias, & Noël, 1998; Lonnemann & Yan, 2015; Moeller et al., 2015; Nuerk, Weger, & Willmes, 2005).

Importantly, cultural differences in numerical processing due to the inversion of number words were not only observed in case the input or output was verbal but were also reported in non-linguistic tasks such as, for instance, the comparison of the magnitude of two numbers presented in Arabic notation (Moeller et al., 2015; Nuerk et al., 2005). Such findings indicate that verbal number word information is co-activated when processing Arabic digits, even when it is not necessary for effective problem solving. If it is indeed co-activated verbal information that leads to the observed differences between language groups, then suppressing the processing of verbal information associated with the numerical task should eliminate differences in numerical processing between language groups. Thus, the primary aim of the current studies was to investigate the effect of articulatory suppression on language differences in number processing.

For different numerical and arithmetic tasks, it has been demonstrated that verbal information influences numerical processing (symbolic magnitude comparison: Moeller et al., 2015; Nuerk et al., 2005; number line estimation: Helmreich et al., 2011; mental addition: Brysbaert et al., 1998; Göbel et al., 2014; Lonnemann & Yan, 2015, Van Rinsveld, Brunner, Landerl, Schiltz, & Ugen, 2015). With respect to two-digit number magnitude comparison, Nuerk and colleagues (2005) investigated verbal influences by using the unit-decade compatibility effect (Nuerk, Weger, & Willmes, 2001): A number pair is considered unit-decade-compatible when comparing tens and units leads to similar response biases (e.g., 42_57: 4 < 5 and 2 < 7). Contrarily, a number pair is unit-decade-incompatible when the respective comparisons lead to opposite response biases (e.g., 47_62: 4 < 6 but 7 > 2). When overall distance is held constant, compatible number pairs are usually responded to faster and with fewer errors than incompatible number pairs. This compatibility effect indicates that the single digits of a two-digit number (i.e., tens and units) are processed separately contradicting the view of holistic processing of the overall magnitude of two-digit numbers in a number magnitude comparison task as argued previously (e.g., Dehaene, Dupoux, & Mehler, 1990; Restle, 1970).

Nuerk and colleagues (2005) showed that, although present in both a German- and an English-speaking sample, the compatibility effect was more pronounced for the German-speaking group. Importantly, this was the case when numbers were presented in Arabic notation. The authors explain this finding by differences in numerical processing due to differences in the verbal number word systems. While the English number word system is fairly consistent for numbers ≥20, German two-digit number words are inverted with respect to the Arabic digit notation (42 is spoken *zweiundvierzig* which literally translates to *two-and-forty*). That means, in German the first named digit for double-digit numbers is the unit digit. Concurrent co-activation of verbal number words might thus have led to increased unit based interference in German, in turn leading to a comparably larger unit-decade compatibility effect for the German sample. Replicating and extending the results by Nuerk et al. (2005), Moeller et al. (2015) demonstrated that both the number word system and reading direction influence the comparison process. The authors showed that the unit-decade compatibility effect is larger when number words and reading direction are incongruent (e.g., German: inverted number words and reading from left to right; Hebrew: not inverted and reading from right to left). The results of Moeller et al. (2015) thus further corroborate the notion that verbal information influences the comparison process and indicate that additional cultural factors have an impact as well.

As regards number magnitude comparison, not only an inverted number word structure but also specific characteristics of the stimulus set were shown to influence the comparison process. For instance, a certain percentage of within-decade filler items (e.g., 35_39) is usually included in stimulus set to prevent participants from focusing on the tens-digit only. Indicating influences of cognitive control (e.g., the adaptation to stimulus characteristics to minimize conflicting information) in basic numerical processing, previous studies revealed that the percentage of within-decade filler items included in the stimulus set influenced the size of the compatibility effects. In this regard, compatibility effects were found to be larger the more within-decade fillers are included in the stimulus set (e.g., Macizo & Herrera, 2011; see Huber, Nuerk, Willmes, & Moeller, 2016, for a computational modeling evidence). Thereby, it is assumed that over the course of an experiment the relative importance of tens and units is modulated depending on the percentage of filler items used. Crucially, influences of an inverted number word structure and specific characteristics of the stimulus set have so far been investigated in isolation. Thus, it is unclear whether language has a differential impact on the comparison process depending on the relative importance of tens and unit digits in a specific stimulus set.

Assuming that the verbal processing component led to the previously observed inversion-related differences in cross-cultural study designs, interfering with the possibility to process verbal information should eliminate the influence of differing number word systems during a numerical task. As a consequence, regular compatibility effects should remain in number magnitude comparison, however, inversion-related differences in compatibility effects between language groups should no longer be present. In this vein, articulatory suppression paradigms are used to investigate the involvement of the articulatory loop (Baddeley, Lewis, & Vallar, 1984). The articulatory loop is part of the verbal working memory and responsible for admitting to and temporarily storing verbal information in the phonological loop (Baddeley, 1992). In articulatory suppression paradigms, participants are usually asked to repeatedly utter a word or a sequence of syllables in a constant rhythm (Baddeley et al., 1984; Murray, 1968). This impairs subvocal repetition processes related to the primary task, thereby preventing admission or storage of additional verbal information. Articulatory suppression paradigms have already been applied in a wide range of tasks and have been used to investigate the involvement of verbal information processing in numerical cognition as well (e.g., Lee & Kang, 2002; Moeller, Klein, Fischer, Nuerk, & Willmes, 2011).

Objectives

Previous studies investigating cross-cultural differences related to number word inversion mostly used quasi-experimental designs (i.e., comparing different language groups). In two studies complementing existing research, we aimed at evaluating linguistic influences on numerical processing as well as their possible interaction with cognitive control processes (e.g., the influence of characteristics of the stimulus set in terms of the percentage of filler items included) more directly. To do so, in Study 1 we manipulated verbal processing resources by means of articulatory suppression during numerical processing *within* two groups of participants with different language backgrounds [i.e., inverted (German) or non-inverted number words (English)]. Moreover, in Study 2, we manipulated the percentage of within-decade filler items included in the stimulus set in addition to the presence/absence of articulatory suppression in two additional samples of, again, German- and English-speaking participants. This allowed us to further evaluate whether influences on the separate processing of tens and units associated with verbal number word processing (i.e., more prominent processing of units in a language with inverted number words) interacts with influences on the separate processing of tens and units associated with stimulus set characteristics (i.e., more prominent processing of units when more within-decade filler items in the set).

Study 1

In Study 1 we investigated verbal processing components in a two-digit number magnitude comparison task using Arabic digits by directly manipulating verbal processing capacities through articulatory suppression. In a mixed model design, German- and English-speaking participants were asked to solve a number magnitude comparison task twice – with and without articulatory suppression. In line with previous results by Nuerk et al. (2005) and Moeller et al. (2015), the unit-decade compatibility effect is expected to be larger for German-speaking (inverted number words) as compared to English-speaking participants (no inversion) when no articulatory suppression is present. However, while regular compatibility effects should still be observed under articulatory suppression, co-activation of verbal number word information is hindered and should, therefore, eliminate differences in the unit-decade compatibility effect between language groups.

Methods

Participants

24 German-speaking participants [M_{age} = 22.25 years, SD_{age} = 4.55, range 18-38, 7 male] were tested at the University of Tuebingen, Germany, and 24 English-speaking participants [M_{age} = 20.54 years, SD_{age} = 3.04, range 18-30, 4 male] were tested at the University of York, UK. The two groups did not differ in age, $t(46) = 1.53$, $p = .133$. All participants were native speakers of the respective language, right-handed and had normal or corrected to normal vision. The study was approved by the ethics committee of the Department of Psychology, University of York (UK).

Stimuli and Design

Participants completed a two-digit number magnitude comparison task. They had to choose the larger of two numbers presented simultaneously and above each other in the centre of a screen. The stimulus set consisted of 480 two-digit number pairs between 21 and 98 in Arabic notation, with 120 unit-decade compatible (e.g., 32_47), 120 unit-decade incompatible (e.g., 37_52; taken from Nuerk et al., 2001), and 240 within-decade filler pairs (e.g., 32_39). Between-decade number pairs realized a 2 x 2 x 2 within-participant design, with the factors compatibility (compatible vs. incompatible), decade distance (small: 1-3 vs. large: 4-8), and unit distance (small: 1-3 vs. large: 4-8) manipulated. Overall distance, decade distance, unit distance, and problem size (i.e., the sum of the two to-be- compared numbers) were matched between the respective stimulus groups.

Procedure

The experiment was presented on an 18" screen with a resolution of 1024 x 768 pixels and a refresh rate of 60 Hz. Participants sat approximately 50 cm in front of the screen. They were instructed to respond as quickly and accurately as possible with the right index finger ("U" key) in case the upper number was larger, and with the left index ("N" key) finger when the bottom number was larger. In half on the cases, the upper number was larger. In the other half, the lower number was larger. Participants had to press the same button not more than twice in a row. Stimuli were presented in white (font: *Courier New*, bold; font size: 24) against a black background until a response was given. Preceding each trial, a fixation cross was presented in the middle of the screen for 500 ms. Trials were separated by an interstimulus interval of 500 ms.

The same design was used for both conditions, with and without articulatory suppression. All participants performed both conditions within one test session. Participants always started with the condition without articulatory suppression to avoid spill-over effects. For both conditions, the experiment started with 12 practice trials, followed by 6 experimental blocks of 80 number pairs each. The experiment lasted approximately 40 min. Trial order was randomized separately for each participant. In the articulatory suppression condition participants had to complete the same magnitude comparison task while uttering the non-sense syllable string "pa-ta-ka" in a staccato fashion at a rate of approximately one syllable per second. The experimenter checked the pace with a stopwatch to ensure loading of the articulatory loop.

Results

Overall error rates were generally low [German: *M* = 4.9%, *SD* = 3.4%; English: *M* = 6.1%, *SD* = 7.1%]. Therefore, analyses focused on RTs of correctly solved trials. A trimming procedure excluded RTs shorter than 200 ms and longer than 3000 ms as well as RTs below or above 3 *SD* of a participant's mean. Trimming resulted in average loss of 1.5% of data [SD =0.5%] for the German-speaking and 1.8% [*SD* = 2.0%] for the English-speaking group. Because German-speaking participants showed slower response times and higher variability in their response times in the condition without articulatory suppression [$t(39.57)=2.11$, $p = .041$; Levene-Test for homogeneity of variances: $p = .022$; M_{Ger}=776ms, SD_{Ger}=131ms; M_{Eng}=710ms, SD_{Eng}=88ms], we z-standardized raw RTs to control for inter-individual differences in raw RT and its variability. For this z-standardization, mean RT and its corresponding SD of all correctly answered items were calculated for each participant and articulatory suppression condition individually and used for standardization. Consequently, possible differences (or the lack thereof) in compatibility effects are not driven by interindividual differences in overall RT (see, e.g., van den Boer, de Jong, & Haentjens-van Meeteren., 2012; Moeller et al., 2015 for a similar approach). Results are given in both zRT and plain RT to ensure readability (see Table 1 for all means (zRT, RT) and standard deviations of all conditions).

< Please insert Table 1 around here >

A 2 × 2 × 2 repeated-measures ANOVA with zRT as dependent variable and the factors compatibility (compatible vs. incompatible), articulatory suppression (with vs. without articulatory suppression), and language group (German vs. English) was conducted. In line with our predictions, the three-way interaction of compatibility, articulatory suppression and language group was significant [$F(1,46) = 6.24$, $p = .016$, $\eta_p^2 = .12$].

Breaking down this three-way interaction into its constituting two-way interactions for the conditions with and without articulatory suppression, respectively, indicated that significant differences in compatibility effects between language groups were present in the condition without articulatory suppression [$F(1, 46) = 4.15$, $p = .047$, $\eta_p^2 = 0.08$]. In contrast to previous observations, however, the observed interaction indicated that the compatibility effect was smaller for German-speaking [$M = 0.26$] than for English participants [$M = 0.36$, see Figure 1]. Moreover, the difference in compatibility effects between language groups was not significant under articulatory suppression [$M_{German} = 0.29$, $M_{English} = 0.26$, $F(1, 46) < 1$, $p = .613$]. To substantiate the null effect observed in the articulatory suppression condition, a Bayesian analysis approach as described by Masson (2011) was applied to allow for the evaluation of the probability of the null hypothesis being true (i.e. that there is no difference in compatibility effects between language groups). Bayesian analyses revealed that the probability of null hypothesis was $p_{BIC} = .858$. According to criteria suggested by Masson (see also Raftery, 1995), this reflects positive evidence for the null hypothesis.

Next to the highest order interaction, the ANOVA revealed that compatible number pairs were on average responded to faster than incompatible number pairs as indicated by the significant main effect of compatibility [$M_{comp} = -0.25$ (740ms), $SD_{comp} = 0.12$, $M_{incomp} = 0.04$ (792ms), $SD_{incomp} = 0.14$; $F(1,46) = 190.80$, $p < .001$, $\eta_p^2 = .81$]. Moreover, responses were faster in the condition without compared to the condition with articulatory suppression [$M_{without} = -0.14$ (743ms), $SD_{without} = 0.20$, $M_{with} = -0.07$ (789ms), $SD_{with} = 0.18$; $F(1,46) = 47.27$, $p < .001$, $\eta_p^2 = .51$]. No other main effects or interactions were significant.

< Please insert Figure 1 around here >

Discussion

As expected, significant compatibility effects were observed indicating reliable unit-based interference in two-digit number processing across language groups caused by separate processing of tens and units. However, the interaction of compatibility, articulatory suppression, and language group was significant indicating that compatibility effects differed between language groups in the condition without articulatory suppression. This between-group difference was smaller and not significant when participants' articulatory loop was blocked by articulatory suppression. Thus, differences in the compatibility effect diminished when co-activation of number words is prevented.

Unexpectedly, however, without articulatory suppression, the compatibility effect was larger in the English- compared to the German-speaking sample. This is not in line with results observed by Nuerk et al. (2005) and Moeller et al. (2015), showing a larger compatibility effect for the German-speaking sample. Most importantly, this combined result pattern cannot be explained solely by the inversion property of German number words. If so, the present data should have revealed a larger compatibility effect for the German-speaking sample indicating increased unit-based interference due to the inverted number word formation in German. Therefore, although between-group differences in compatibility effects disappeared under articulatory suppression, the observation of a reversed between-group difference in the no articulatory suppression condition (i.e., a larger compatibility effect for English-speaking participants) may question the assumption that specificities in the number word structure only led to the observed between-group differences. As such, these results do not yet allow for a definitive answer onto what is actually eliminated in the articulatory suppression condition.

One possible explanation for the observed difference between studies may lie in the composition of the stimulus set used or more precisely, in the percentage of within-decade filler items included in the set. While Nuerk et al. (2005) used no filler items and Moeller et al. (2015) used 33% filler items, 50% filler items were included in the stimulus set used in the present study. When filler items indeed modulate the relevance of tens and units, it is possible that language affects the comparison process differentially depending on the correspondence of the respective number word structure and the relevance of tens and units in a given stimulus set.

Without any filler items, the unit digit does not have to be considered at all, because the tens digit is sufficient to make the correct decision in 100% of all cases. When 33% of the trials are fillers, the correct decision can be derived in 67% of the trials from either the tens (in the 67% critical between-decade pairs) or the unit digit (in the 33% compatible between-decade pairs and 33% within-decade fillers). However, here the unit digit only has to be considered actively in the minority of trials (i.e., in the 33% filler pairs). Thus, the tens digit is still more relevant. In this case, speaking a non-inverted language may be advantageous because naming the tens digit, which is of higher decision relevance, first, might reduce the interference caused by decision-irrelevant unit digit in critical trials. In turn, this should lead to a smaller compatibility effect in speakers of a language with non-inverted number words (as observed in Moeller et al., 2015; Nuerk et al., 2005).

The case is, however, different when the stimulus set contains 50% of filler items - as in Study 1. Focusing on the tens digit leads to the correct decision in only half of the trials (50% critical between-decade trials). In contrast, focusing on the unit digit allows for deriving the correct decision in 75% of all cases (50% filler and 25% compatible between-decade trials). Thus, with 50% fillers included in the stimulus set, primarily considering the tens digit is no longer a superior strategy. Rather, focusing on the unit digit now seems to be a more economic strategy because in the majority of cases (75%) comparing the unit digit leads to the correct decision (see Huber, Mann, Nuerk, & Moeller, 2013 for eye-tracking evidence corroborating this argument). In turn, when a focus on the unit digits is beneficial for the stimulus set at hand, speaking a language with inverted number words might be advantageous. As the unit digit is named first in inverted number words, such a setting may thereby facilitate the comparison process. Based on past (Moeller et al., 2015; Nuerk et al., 2005) and present results, it is, therefore, suggested that number word inversion has a differential impact on the comparison process depending on the relative importance of tens and unit digits. This importance is influenced by the percentage of filler items used in the study.

Study 2

To directly test influences of the percentage of filler items on language differences in compatibility effects, Study 2 was conducted. In addition to manipulating language (German vs. English) and articulatory suppression (with vs. without articulatory suppression), in Study 2 we also manipulated the percentage of filler items within participants (50% vs. 20% filler items). Thereby, Study 2 allowed for a direct replication of Study 1 with respect to the 50% filler condition. Furthermore, it allowed us to test the hypothesis of differential compatibility effects depending on both language and the percentage of filler items in the stimulus set as described above.

First, we hypothesized to replicate the results of Study 1. In particular, we expected a larger compatibility effect for English- compared to German-speaking participants when 50% within-decade filler items are included in the set. Additionally, we hypothesized that this between language group difference disappears under articulatory suppression. However, in the new 20% filler condition, we expected a larger compatibility effect for German- as compared to English-speaking participants, in line with previous studies using 33% or less filler items (e.g., Moller et al., 2015; Nuerk et al., 2005). Additionally, this between-group difference should disappear under articulatory suppression as well.

Methods

Participants

For Study 2, new participants were recruited. 24 German-speaking participants (M_{age} = 24.58 years, SD_{age} = 3.89, range 19-34, 8 male) were tested at the University of Tuebingen, Germany, and 24 English-speaking participants at the University of York, UK. One of the English-speaking participants was excluded because data were only available for two out of four experimental conditions. Mean age of the remaining English-speaking sample was M_{age} = 19.64 years (SD_{age} = 2.44, range 18-30, 4 male). German-speaking participants were on average older than their English-speaking counterparts [$t(39.11) = 5.21$, $p < .001$]. All participants were native speakers of the respective language, right-handed and had normal or corrected to normal vision. The study was approved by the ethics committee of the Department of Psychology, University of York (UK).

Stimuli and Design

Participants of both language groups completed a total of four blocks of number magnitude comparison because in addition to the manipulation of articulatory suppression (with vs. without) the percentage of filler items was manipulated within-subject (50% vs. 20% filler items). The 50% filler conditions were identical to the ones used in Study 1. For the 20% filler conditions, the stimulus set of Study 1 was reduced so that it contained the same 240 experimental items but only 60 within-decade filler items.

Procedure

The procedure of Study 2 was identical to that of Study 1 except for the fact that participants had to come to the lab twice with at least one day (24h) between the two sessions. In the first session, they performed both blocks of one filler condition. Order of filler conditions was balanced across participants so that half of the participants started with the 20% filler conditions and the other half started with the 50% filler conditions in the first session. Participants always performed the block without concurrent articulation first regardless of filler condition. Each block started with 12 practice trials, followed by six experimental runs of 80 number pairs each in the 50% filler condition and five runs of 60 number pairs in the 20% filler condition. Instructions, stimulus presentation as well as response keys were the same as in Study 1. For the 50% filler condition, the experiment lasted approximately 40 min, with about 20 min for each condition. For the 20% filler condition, the experiment lasted approximately 30 min, with about 15 min for each condition.

Results

Analyses steps were identical for Study 1 and 2. Error rates were low [German: M = 5.4%, SD = 4.1%; English: M = 5.8%, SD = 4.1%]. Trimming resulted in average loss of 1.4% of data [SD =0.3%] for the German-speaking and 1.3% [SD = 0.3%] for the English-speaking group. All means (zRT, RT) and standard deviations of all conditions are provided in Table 1.

A 2 × 2 × 2 × 2 repeated-measures ANOVA with zRT as dependent variable discerning the factors compatibility (compatible vs. incompatible), articulatory suppression (with vs. without), language group (German vs. English), and percentage of fillers (50% vs. 20%) was conducted to directly tests the hypothesis of differential effects of number word formation on the compatibility effect depending of the percentage of within-decade filler items. In line with our predictions, the four-way interaction of compatibility, articulatory suppression, percentage of filler items and language group was significant [$F(1,45) = 7.25$, $p = .010$, $\eta_p^2 = .14$]. Breaking down this four-way interaction into its constituting three-way interactions for the condition without and with articulatory suppression, respectively, showed that the three-way interaction was only significant in the condition without articulatory suppression [$F(1,45) = 10.71$, $p = .002$, $\eta_p^2 = .19$] but not in the condition with articulatory suppression [$F < 1$, $p = .873$]. Bayesian analyses substantiated this null effect by revealing that the posterior probability of the null hypothesis was $p_{BIC} = .871$. According to criteria suggested by Masson (2011; see also Raftery, 1995), this reflects positive evidence for the null hypothesis. Further breaking down the significant three-way interaction observed in the condition without articulatory suppression showed that when 50% fillers were included in the stimulus set – as it was the case in Study 1 - the compatibility effect tended to be larger for the English- as compared to the German-speaking sample [$M_{German} = 0.28$, $M_{English} = 0.37$, $F(1, 45) = 3.45$, $p = .070$, $\eta_p^2 = 0.07$; see Figure 1]. However, and in line with our predictions, in case 20% fillers were included in the stimulus set the compatibility effect was larger for the German- as compared to the English-speaking sample [$M_{German} = 0.29$, $M_{English} = 0.20$, $F(1,45) = 5.73$, $p = .021$, $\eta_p^2 = .11$, see Figure 2].

Next to this highest order interaction, the ANOVA revealed that compatible number pairs were on average responded to faster than incompatible number pairs as indicated by the significant main effect of compatibility [$M_{comp} = -0.13$ (692ms), $SD_{comp} = 0.30$, $M_{incomp} = 0.15$ (735ms), $SD_{incomp} = 0.30$; $F(1,45) = 300.17$, $p < .001$, $\eta_p^2 = .87$]. Additionally, reaction times in the 20% filler condition were faster compared to the 50% filler condition [$M_{50\%} = 0.10$ (726ms) , $SD_{50\%} = 0.33$, $M_{20\%} = -0.08$ (701ms), $SD_{20\%} = 0.31$; $F(1,45) = 6.73$, $p = .013$, $\eta_p^2 = .13$]. Additionally, overall reaction times of the German speakers were faster compared to English speakers [$M_{German} = 0.007$ (688ms), $SD_{German} = 0.34$, $M_{English} = 0.014$ (740ms) , $SD_{English} = 0.33$; $F(1,45) = 6.13$, $p = .017$, $\eta_p^2 = .12$]. Finally, and in line with previous studies (e.g., Macizo & Herrera, 2011), the compatibility effect was more pronounced when 50% fillers were included in the stimulus set [$M_{50\%} = 0.32$ (52ms), $SD_{50\%} = 0.18$, $M_{20\%} = 0.22$ (33ms) , $SD_{20\%} = 0.15$; $F(1,45) = 20.87$, $p < .001$, $\eta_p^2 = .32$]. No other main effects or interactions were significant.

< Please insert Figure 2 around here >

Discussion

To replicate the results of Study 1 and to evaluate the unexpected result of a larger compatibility effect in English- as compared to German-speaking participants, the percentage of filler items included in the stimulus set was manipulated in addition to language and articulatory suppression in Study 2. Results were in line with our expectations. First, we replicated the effect of articulatory suppression already observed in Study 1: Differences in the compatibility effect between language groups disappeared in conditions with articulatory suppression. Second, we replicated the unexpected result pattern of Study 1: In the 50% filler condition in Study 2 we again observed that the compatibility effect tended to be larger for English-speaking as compared to German-speaking participants.

Finally, results of Study 2 corroborated our hypothesis of differential language influences on the compatibility effect depending on the percentage of filler items included in the stimulus set. In line with previous observations of studies including 33% or less filler items (Moeller et al., 2015; Nuerk et al., 2005), we found that the compatibility effect was larger in the German- compared to the English-speaking sample in the 20% filler condition of Study 2. However, this reversed in the 50% filler condition. For the 20% filler condition in Study 2, focusing on the *tens* digit allowed for deriving the correct decision in the majority of cases (80% compatible between-decade trials). Here, speaking a non-inverted language (tens digit is named first) is consistent with the relevance of tens and units determined by the characteristics of the stimulus set which seems to have led to the smaller compatibility effect for English- as compared to German-speaking participants. Contrarily, in case 50% fillers were included in the set (Study 1 and Study2), focusing on the *unit* digit allowed for deriving the correct decision in the majority of cases (75%: 50% filler and 25% compatible between-decade trials). Thus, when considering characteristics of the stimulus set, focusing on the unit digit was a more economic strategy. In this case, speaking a language with inverted number words (unit digit is named first) seems to be advantageous resulting in smaller compatibility effects for German- compared to English-speaking participants.

Considering the overall pattern of results, Study 2 provides two additional new insights. First, verbal (number word) information seems to be activated during symbolic Arabic number magnitude comparison irrespective of the percentage of fillers included in the stimulus set because between-language group effects disappeared under articulatory suppression in both filler conditions. Second, the numerical comparison process seems to be most efficient in case the number word structure of a given language corresponds to the relative importance of tens and units as determined by characteristics of the stimulus set. In particular, Study 2 suggests that naming the unit digit first (as it is the case in languages with inverted number words such as German) does not necessarily lead to worse performance outcomes due to additional interference of the unit digit but may actually be an advantage in case the respective focus on the unit digit corroborates a more efficient solution strategy (as is the case when 50% within-decade fillers are included in the stimulus set).

General Discussion

In two studies we directly evaluated the underlying mechanisms of previously observed linguistic influences on symbolic Arabic two-digit number magnitude comparison using an articulatory suppression paradigm. In Study 1, German- and English-speaking participants performed a number magnitude comparison task twice: with and without articulatory suppression (e.g., uttering pa-ta-ka during the comparison process). In Study 2, we additionally manipulated the percentage of filler items included in the stimulus set. Results of both studies consistently support the notion that – although not necessary for the comparison process – number words are automatically co-activated in number magnitude processing even in a nonverbal numerical task using Arabic digits. Specifically, results show that it is indeed concurrent processing of number words that leads to the observed differences between language groups because between-language group differences in compatibility effects disappeared under articulatory suppression. Study 2 further indicated that the inversion of number words does not per se lead to worse performance in the comparison process due to additional interference of the first named unit digit. Instead, the direction of between-language group differences depended on the percentage of fillers included in the stimulus set. In particular, efficiency in comparing two two-digit numbers seemed to depend on the correspondence of the number word structure with the relevance of tens and units which is determined by the overall characteristics of the stimulus set.

This differential influence of articulatory suppression in two-digit number magnitude comparison expands previous findings of Nuerk et al. (2005) and Moeller et al. (2015) by applying a more direct test of co-activation of verbal numerical information whilst solving the task. Most importantly, in both Study 1 and Study 2 and, thus, irrespective of the percentage of filler items included in the set, between group differences in the compatibility effect disappeared when articulatory suppression was applied. Results are, therefore, not in line with the assumption that number magnitude comparison tasks presented in Arabic notation are processed without any involvement of the verbal code as proposed by the currently most influential model of number processing (e.g., Dehaene & Cohen, 1995; Dehaene, Piazza, Pinel, & Cohen, 2003).

However, in contrast to previous findings, differential between-language group patterns were observed depending on the percentage of filler items included in the stimulus set. This further indicates that the inversion of number words might not be the only factor driving differences in two-digit number processing between language groups. Instead, these new findings suggest that linguistic differences and properties of the stimulus set (e.g., the percentage of filler items used) interact in symbolic number processing. Therefore, when aiming to determine the underlying processes of two-digit number magnitude comparison it is crucial to consider both cross-linguistic differences and influences of cognitive control (driven by stimulus set characteristics) as well as their interaction.

Acknowledgements

Julia Bahnmueller was supported by the Leibniz-Competition Fund (SAW-2014-IWM-4). We would like to thank Charlotte Ashton, Lia Heubner, and Marie-Lene Schlenker, for help with data collection.

References

Baddeley, A. (1992). Working memory. *Science, 255(5044),* 556-559. http://doi.org/10.1126/science.1736359

Baddeley, A., Lewis, V., & Vallar, G. (1984). Exploring the articulatory loop. *The Quarterly Journal of Experimental Psychology, 36(2),* 233-252. http://doi.org/10.1080/14640748408402157

Brysbaert, M., Fias, W., & Noël, M.-P. (1998). The Whorfian hypothesis and numerical cognition: Is “twenty-four” processed in the same way as “four-and-twenty”? *Cognition*, *66*(1), 51–77. http://doi.org/10.1016/S0010-0277(98)00006-7

Comrie, B. (2005). Endangered numeral systems. In J. Wohlgemuth & T. Dirksmeyer (Eds.), *Bedrohte Vielfalt: Aspekte des Sprach(en)tods* [Endangered Diversity: Aspects of Language Death] (pp. 203-230). Berlin: Weißensee.

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

Dehaene, S., Dupoux, E., & Mehler, J. (1990). Is numerical comparison digital? Analogical and symbolic effects in two-digit number comparison. *Journal of Experimental Psychology: Human Perception and Performance*, *16*(3), 626–641. http://doi.org/10.1037/0096-1523.16.3.626

Dehaene, S., Piazza, M., Pinel, P., & Cohen, L. (2003). Three parietal circuits for number processing. *Cognitive Neuropsychology*, 20, 487–506. http://doi.org/10.1080/02643290244000239

Göbel, S., 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. http://doi.org/10.1016/j.jecp.2013.10.001

Helmreich, I., Zuber, J., Pixner, S., Kaufmann, L., Nuerk, H.-C., & 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. *Journal of Cross-Cultural Psychology, 42*, 598-613. http://doi.org/10.1177/0022022111406026

Huber, S., Mann, A., Nuerk, H.-C., & Moeller, K. (2013). Cognitive control in number magnitude processing - Evidence from eye-tracking. *Psychological Research, 78*, 539-548. http://doi.org/10.1007/s00426-013-0504-x

Huber, S., Nuerk, H., Willmes, K., & Moeller, K. (2016). A general model framework for multisymbol number comparison. *Psychological Review*, *123*(6), 667–695. http://doi.org/10.1037/rev0000040

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*, *5:313*. http://doi.org/10.3389/fpsyg.2014.00313

Lee, K. M., & Kang, S. Y. (2002). Arithmetic operation and working memory: Differential suppression in dual tasks. *Cognition, 83*, B63-B68. http://doi.org/10.1016/S0010-0277(02)00010-0

Lonnemann, J., & Yan, S. (2015). Does number word inversion affect arithmetic processes in adults? *Trends in Neuroscience and Education*, *4*(1–2), 1–5. http://doi.org/10.1016/j.tine.2015.01.002

Macizo, P., & Herrera, A. (2011). Cognitive control in number processing: Evidence from the unit–decade compatibility effect. *Acta Psychologica*, *136*(1), 112–118. http://doi.org/10.1016/j.actpsy.2010.10.008

Masson, M. E. J. (2011). A tutorial on a practical Bayesian alternative to null-hypothesis significance testing. *Behavior Research Methods, 43*, 679-690. http://doi.org/10.3758/s13428-010-0049-5

Miller, K. F., Kelly, M., & Zhou, X. (2005). Learning mathematics in China and the United States: Cross-cultural insights into the nature and course of preschool mathematical development. In J. I. D. Campbel (Ed.), *Handbook of Mathematical Cognition* (pp. 163–178). New York: Psychology Press.

Miura, I. T., & Okamoto, Y. (2003). Language supports for mathematics understanding and performance. In A. J. Baroody & A. Dowker (Eds.), *The Development of Arithmetic Concepts and Skills: Constructing Adaptive Expertise. Studies in Mathematics Thinking and Learning* (pp. 229–242). Mahwah, NJ: Erlbaum.

Moeller, K., Klein, E., Fischer, M. H., Nuerk, H.-C., & Willmes, K. (2011). Representation of multiplication facts – Evidence for partial verbal coding. *Behavioral and Brain Functions, 7:25*. http://doi.org/10.1186/1744-9081-7-25

Moeller, K., Shaki, S., Göbel, S. M., & Nuerk, H.-C. (2015). Language influences number processing – A quadrilingual study. *Cognition, 136*, 150-155. http://doi.org/10.1016/j.cognition.2014.11.003

Murray, D. J. (1968). Articulation and acoustic confusability in short-term memory. *Journal of Experimental Psychology, 78,* 679-684. http://doi.org/10.1037/h0026641

Ngan Ng, S. S., & Rao, N. (2010). Chinese number words, culture, and mathematics learning. *Review of Educational Research*, *80*(2), 180–206. http://doi.org/10.3102/0034654310364764

Nuerk, H.-C., Weger, U., & Willmes, K. (2001). Decade breaks in the mental number line? Putting the tens and units back in different bins. *Cognition, 82*(1), B25-B33. http://doi.org/10.1016/S0010-0277(01)00142-1

Nuerk, H.-C., Weger, U., & Willmes, K. (2005). Language effects in magnitude comparison: Small, but not irrelevant. *Brain and Language, 92*(3), 262-277. http://doi.org/10.1016/j.bandl.2004.06.107

OECD (2014). PISA 2012 Results: What Students Know and Can Do – Student Performance in Mathematics, Reading and Science (Volume I, Revised edition, February 2014). PISA, OECD Publishing.

Okamoto, Y (2015). Mathematics learning in the USA and JAPAN: Influences of language. In R. Cohen Kadosh & A. Dowker (Eds.), *Handbook of Mathematical Cognition* (pp. 415-429). Oxford University Press, Oxford.

Raftery, A. E. (1995). Bayesian model selection in social research. In P. V. Marsden (Ed.), *Sociological Methodology 1995,* 111-196. Cambridge: Blackwell.

Restle, F. (1970). Speed of adding and comparing numbers. *Journal of Experimental Psychology*, *83*(2, Pt.1), 274–278. http://doi.org/10.1037/h0028573

van den Boer, M., de Jong, P. F., Haentjens-van Meeteren, M. M. (2012). Lexical decision in children: Sublexical processing or lexical search? *The Quarterly Journal of Experimental Psychology*, *65*, 1214–1228. http://doi.org/10.1080/17470218.2011.652136

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

Xenidou-Dervou, I., Gilmore, C., van der Schoot, M., & van Lieshout, E. C. (2015). The developmental onset of symbolic approximation: Beyond nonsymbolic representations, the language of numbers matters. *Frontiers in Psychology*, *6:487*. http://doi.org/10.3389/fpsyg.2015.00487

Table 1. Mean zRT and standard deviations as well as mean RT for all conditions of Study 1 and Study 2.

	language	articulatory suppression	compatibility	zRT	SD	RT (ms)
STUDY 1	*German*	*with*	*compatible*	-0.23	0.12	739
			incompatible	0.06	0.13	791
		without	*compatible*	-0.28	0.12	754
			incompatible	-0.02	0.15	799
	English	*with*	*compatible*	-0.19	0.11	785
			incompatible	0.08	0.13	842
		without	*compatible*	-0.31	0.10	684
			incompatible	0.05	0.15	735

	language	articulatory suppression	filler	compatibility	zRT	SD	RT (ms)
STUDY 2	*German*	*with*	*50*	*compatible*	-0.01	0.31	691
				incompatible	0.33	0.28	743
			20	*compatible*	-0.25	0.28	662
				incompatible	-0.04	0.34	691
		without	*50*	*compatible*	-0.00	0.27	675
				incompatible	0.27	0.20	716
			20	*compatible*	-0.27	0.24	641
				incompatible	0.02	0.25	682
	English	*with*	*50*	*compatible*	-0.14	0.36	734
				incompatible	0.16	0.31	793
			20	*compatible*	-0.07	0.32	750
				incompatible	0.10	0.32	782
		without	*50*	*compatible*	-0.08	0.28	700
				incompatible	0.29	0.24	756
			20	*compatible*	-0.17	0.27	688
				incompatible	0.02	0.22	718

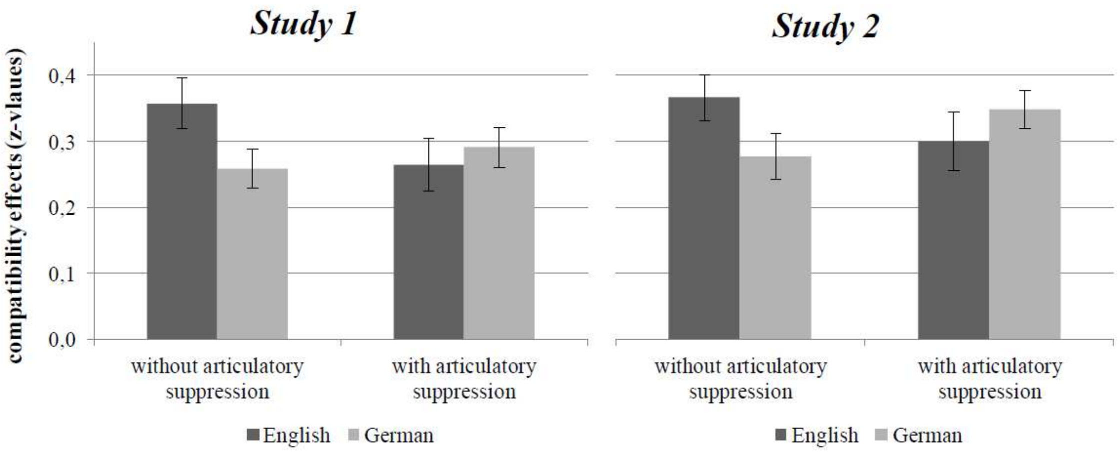

Figure 1. Compatibility effects for German- and English-speaking participants without and with articulatory suppression presented separately for Study 1 and Study 2. Presented are 50% filler conditions only. Error bars indicate 1 standard error of the mean (SEM).

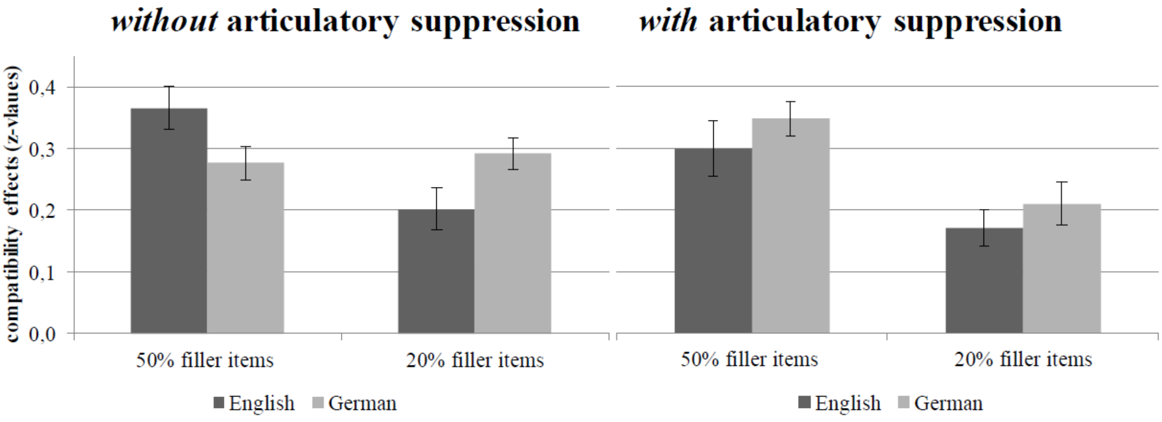

Figure 2. Compatibility effects for the two language groups (German vs. English) and the two filler conditions (50% vs. 20% fillers) presented separately for the condition without and with articulatory suppression, respectively. Error bars indicate 1 standard error of the mean (SEM).