Syntactic priming reveals an explicit syntactic representation of multi-digit verbal numbers[°]

Dror Dotan, Ilya Breslavskiy, Haneen Diab-Copty, and Vivian Yousefi

Mathematical Thinking Lab, School of Education and the Sagol School of Neuroscience, Tel Aviv University

When we say or understand verbal numbers, a major challenge to the cognitive system is the need to process the number's syntactic structure. Several studies showed that number syntax is handled by dedicated processes, however, it is still unclear how precisely these processes operate, whether the number's syntactic structure is represented explicitly, and if it is – what this representation looks like. Here, we used a novel experimental paradigm, syntactic priming of numbers, which can examine in detail the syntactic representation of multi-digit verbal numbers. In each trial, the participants – Arabic-Hebrew bilinguals and Hebrew monolinguals – heard a multi-digit number and responded orally with a random number. The syntactic structure of their responses was similar to that of the targets, showing that they represented the verbal number's syntax. This priming effect was genuinely syntactic, and could not be explained as lexical – repeating words from the target; as phonological – responding with words phonologically-similar to the target; or as a numerical distance effect – producing responses numerically close to the target. The syntactic priming effect was stronger for earlier words in the verbal number and weaker for later words, suggesting that the syntactic representation is capped by working-memory limits. We propose that syntactic priming could become a useful method to examine various aspects of the syntactic representation of numbers.

1. Introduction

1.1. The syntax of numbers

Most languages can denote any integer number up to a reasonable size as a sequence of words using a small lexicon of words and a generative system of syntactic rules. In the cognitive system too, there are lexical and syntactic processes to handle numbers: lexical processes handle basic lexical elements, i.e., digits or number words, and syntactic processes merge lexical elements into syntactically-structured multi-digit numbers (or decompose multi-digit numbers into lexical elements). The lexical-syntactic distinction has been demonstrated by neuropsychological studies that showed double dissociations between lexical and syntactic processing: lexical processing can be selectively impaired while sparing syntactic processing (Cohen et al., 1997; Dotan & Friedmann, 2015; McCloskey et al., 1985, 1986); and syntactic processing can be impaired while sparing lexical processing (Cappelletti et al., 2005; Cipolotti, 1995; Cipolotti et al., 1994; Dotan & Friedmann, 2018; Furumoto, 2006; Noël & Seron, 1993).

[°] We thank Daled Dotan and Ricardo Tarrasch for their help in the data analysis. This research was supported by the Jacobs Foundation (grant no. 2019-1320-05, Dotan).

Of these two types of processes – lexical and syntactic – the bigger challenge to the cognitive system seems to be posed by the processing of syntax. Multiple types of evidence indicate this syntactic difficulty. First, learning the syntax of numbers during childhood is hard and takes years to master (Cheung & Ansari, 2020; Dotan & Dehaene, 2016). Second, children make more syntactic errors than lexical errors when reading or writing numbers (Moura et al., 2013; Power & Dal Martello, 1990, 1997; Seron & Fayol, 1994). Even adults, who make fewer errors when they read numbers, still make more syntactic than lexical errors (Dotan & Friedmann, 2018). Third, cognitive deficits in syntactic processes seem to be more common than deficits in lexical processes. We are not familiar with published statistics about the prevalence of specific types of number reading/writing disorders, but when we examined the precise locus of deficit for 23 randomly-selected adults with dysnumeria (number reading impairment), each single one of them had a syntactic deficit, whereas only 5 had lexical deficits (Dotan, unpublished data). For children too, children with mathematical disorders make more syntactic than lexical errors (Moura et al., 2013). This idea, that syntactic processing of numbers is more demanding cognitively than their lexical processing, accords with the view that the ability to process certain syntactic structures is an advanced cognitive ability, which is unique to humans and is not shared by animals (Dehaene et al., 2015; Hauser et al., 2002).

Clearly, number syntax is necessary for transcoding – reading aloud multi-digit strings or multi-word numbers, or writing them to dictation. Nevertheless, the importance of number syntax may go beyond transcoding tasks – it may affect mathematical achievements in general. Of the three cognitive representations of numbers – symbolically as digits or number words, and non-symbolically as approximate quantities (Dehaene, 1992) – recent studies indicate that the ability to process number symbols predicts mathematical performance better than the ability to process quantities (Göbel, 2020; Malone et al., 2021; Moeller et al., 2011; Xenidou-Dervou et al., 2017; Yuan et al., 2019). A major aspect of this ability to process symbolic numbers is the processing of their syntactic structure (Dotan & Dehaene, 2020b; Dotan et al., 2021; Dotan & Friedmann, 2018; McCloskey, 1992). Indeed, children’s understanding of number syntax (place-value principle) predicts later achievements in math (Miura & Okamoto, 1989).

1.2. The cognitive representation of number syntax

Several neuropsychological studies, including those cited above, show that number syntax can be selectively impaired. These studies clearly indicate the existence of cognitive processes dedicated to syntax. However, the precise nature of these cognitive processes is yet unclear.

One possibility is that number syntax is handled by several processes, each handling a specific type of syntax-relevant information. For example, errors that occur specifically in the digit 0 (Furumoto, 2006; Granà et al., 2003), or specifically in non-0 digits (Cohen & Dehaene, 1991; Dotan & Friedmann, 2018), may be attributed to a selective deficit in dedicated processes that handle the digit 0, or to a selective sparing of such processes, either in the visual analysis of digit

strings (Dotan et al., 2021) or in the digit-writing mechanisms (Furumoto, 2006). Errors in the number length (e.g., reading 12 as if it had 3 digits, “one hundred and two”, Dotan & Friedmann, 2018) may arise from a deficit in a specific process that handles the length information. For example, in the visual analysis of digit strings, a number-length detector may help setting up the visual analyzer to handling a particular number of digits (Dotan & Dehaene, 2020b; Dotan et al., 2021). In the verbal system, the dissociation between digit substitutions (e.g., “twenty three” → “twenty seven”) and lexical class substitutions (e.g., “twenty three” → “twenty thirty”) may be explained as a syntactic-semantic organization of the number-word phonological lexicon, e.g., if each number word is represented as two morphemes, a digit and a class (McCloskey et al., 1986). All these syntactic processes may be considered as low-level, because they do not represent the number’s full syntactic structure but only some aspects of it; and because they are peripheral, in the sense that they operate in the context of a particular input or output process, in a particular modality.

Another interesting possibility, which we examine here, is that on top of these low-level syntactic processes, there is an explicit representation of the number’s full syntactic structure. This was one of the central ideas in Michael McCloskey’s number processing model, which we now explain.

1.2.1. McCloskey’s model of number processing

The processing of number syntax can be broadly defined as processing the relations between lexical elements – digits and number words. As a more concrete definition of number syntax, McCloskey and colleagues proposed in the mid-1980s an influential cognitive model (McCloskey, 1992; McCloskey et al., 1986), which we will examine in the present study. According to this model, symbolic numbers (digits and number words) have a syntactic-semantic abstract core representation. The representation is abstract in the sense that it is neither verbal nor visual. It is semantic, as it contains the full information about the number (all digits). Finally, it is syntactic, as it reflects the number’s base-10 structure – e.g., the representation of 37 is denoted in McCloskey’s model as 3EXP2 & 7EXP1, similar to the number’s decimal structure ($37 = 3 \times 10^1 + 7 \times 10^0$). McCloskey also hypothesized that the syntactic representation is hierarchical: in the syntactic representation of a number, the units and decades are merged first, then this pair is merged with the hundreds (thereby forming a triplet), and then two triplets can be merged together. For example, the number 234,567 would be merged as [2 & (3 & 4)] & [5 & (6 & 7)] – a hierarchical tree-like structure. Dotan and Friedmann (Dotan & Friedmann, 2018) proposed a detailed cognitive model of number reading which embraces and extends McCloskey’s assumption of a hierarchical syntactic representation.

McCloskey’s model further assumes that this syntactic-semantic representation is not the whole story of number syntax: additional syntactic processes exist as part of the mechanisms that translate

the central representation into a verbal number or a digit string, or create the central representation from verbal numbers or digit strings.

The present study examined McCloskey's assumption that the full syntax of multi-digit numbers is represented explicitly. We examined this assumption specifically with respect to verbal numbers.

1.2.2. Criticisms on McCloskey's model and the survival of its assumptions about number syntax

During the 1990s there was a lively theoretical debate that contrasted McCloskey's model versus other theoretical models of number processing, the main competition being from Dehaene's triple-code model (Dehaene, 1992; Dehaene & Cohen, 1995). Critical assumptions of McCloskey's model were refuted, in particular the assumption of an abstract semantic representation that serves as an intermediate representation in almost any number processing task, in particular calculation and transcoding. Crucially, however, these refutations did not challenge McCloskey's notion of a *syntactic* representation, but only the notion of a semantic representation; and they refuted the idea of a *mandatory* abstract representation, i.e., one that must be used in any task, but they did not refute the possibility that such representation is used in some tasks but not in others.

The two main criticisms on McCloskey's model were as follows. First, some researchers criticized McCloskey's assumption that calculation operates on a central abstract representation. This assumption predicts that the performance in a calculation task depends on the abstract representation but should be independent of the input and output format – i.e., on the number's presentation mode (verbally, as Arabic digits or digit strings, as Roman numerals, etc.) and on the response model (oral, writing, etc.). Contrary to this prediction, several studies showed that the input/output format did in fact affect the performance in calculation (Campbell & Clark, 1992; González & Kolers, 1982; Noël & Seron, 1997). Nevertheless, while these studies speak against the notion of a central representation that serves all calculation operations, they do not refute the possibility of a representation that is used for operations other than calculation, for example, for reading or writing numbers.

Other researchers criticized McCloskey's assumption that the syntactic-semantic representation must serve as a mediating representation in any transcoding task – reading aloud digit strings or writing numbers to dictation. To refute this idea, two important studies (Cohen & Dehaene, 1991, 2000) examined the performance of patients with brain damage who made many errors when reading numbers aloud. McCloskey's model predicts that such errors must originate in a deficit at the visual input stage (which would affect any comprehension task), at the verbal production stage (which would affect any task in which numbers are produced verbally), or in the central abstract representation (which would affect reading, writing, calculation, and additional tasks). Contrary to this prediction, the patient described by Cohen and Dehaene (2000) had good comprehension

(refuting a visual-input deficit), good production of verbal numbers (refuting a verbal-production deficit), and good performance in written calculation relative to oral calculation (refuting a central-representation deficit); and the patient described by Cohen and Dehaene (1991) had good comprehension (refuting a visual deficit) yet his errors were clearly affected by visual factors (refuting a post-visual deficit). Thus, the performance of these two patients could not be explained by McCloskey's model. Nevertheless, both patients made only digit substitutions errors, with no syntactic errors, so their performance does not inform about the processing of number syntax, but only about semantic processing. Namely, these two cases refute the extreme assumption that a central representation of the full number semantics mediates number reading, but they do not refute a narrower assumption, according to which there is a central *syntactic* representation of the number.

In short, none of the studies above has challenged McCloskey's assumption of a syntactic representation of the multi-digit number; and none has refuted the idea of an "optional" central representation, i.e., one that is used in some tasks but not in others.

1.2.3. Existing evidence for syntactic processing

1.2.3.1. *Syntactic processing of verbal numbers*

Many studies have demonstrated the existence of cognitive processes dedicated to handling number syntax. For example, several neuropsychological case studies showed that the processing of number syntax can be selectively impaired (Cipolotti, 1995; Delazer & Bartha, 2001; Dotan & Friedmann, 2018; Noël & Seron, 1993). However, as explained above, number syntax is a broad concept, and most of the studies that examined syntactic processing did not identify the specific syntactic process involved. Here, our goal was to focus on verbal numbers and to examine McCloskey's specific assumption of an explicit representation of the number syntax. As we shall now review, only a few studies specifically addressed the syntactic representation of verbal numbers, and none showed an explicit representation of the number's syntactic structure.

One study, from our own lab (Dotan & Friedmann, 2018), examined several individuals with developmental dysnumeria (number reading disorder). One of them (OZ) had a selective deficit in creating the *number word frame* – the series of lexical classes (ones, tens, teens, etc.) in a verbal number. OZ's difficulty may be attributed to a deficit in an abstract syntactic representation as the one hypothesized by McCloskey, but it may also be attributed to deficits in other processes – e.g., a verbal, language-specific syntactic process. Furthermore, OZ's pattern showed a syntactic deficit but did not show an explicit representation of the full number syntax. Thus, while OZ's case certainly demonstrates a syntactic sub-process in verbal number production, it does not show an explicit representation of the number's syntactic structure.

In another study (Hung et al., 2015), participants read aloud sequences of number words – grammatical (e.g., Five hundred sixty seven), non-grammatical (seven sixty five and hundred), and

semi-grammatical (sixty seven five and hundred – i.e., the full sequence is not grammatical, but some segments in it are). The participants' reaction times and brain activity patterns were affected by the syntactic complexity of the number word sequence: sequences with more/longer grammatical segments yielded shorter reaction time and higher brain activity in the left IFG and the left IPL. These findings suggest a mechanism that merges grammatical number word pairs (and perhaps even longer segments). Still, Hung et al. did not examine the precise nature of this merge operation. In particular, they did not examine whether this merge involves an explicit representation of the full number syntax (or, for example, operates only on adjacent word pairs).

Other studies used number dictation task with multi-digit numbers and showed that the processing depended on the verbal number's syntactic structure. Lochy et al. (2002) showed that the time gaps between the writing of consecutive digits reflected the syntactic structure of the verbal number the participants heard. Specifically, if a 4-digit number was dictated as *teens, decades, and units* (e.g., 1234 as “twelve hundred thirty-four”), there was a large gap between the hundreds and tens digit; but if the number was dictated as *thousands, hundreds, decades, and units* (one thousand, two hundred and thirty-four), there was a large gap between the thousands and hundreds digits. Noël and Seron (1997) used a similar manipulation: they showed that comparing a pair of 4-digit numbers was faster when the two numbers were presented in the same verbal format (e.g., twelve hundred vs. fourteen hundred) rather than in different formats (twelve hundred vs. one thousand and four hundred); and that in oral 4-digit addition ($XX00 + X00$), participants tended to respond using the verbal format presented in the question. These studies clearly show that the verbal number's syntax was parsed at some stage, however, they not demonstrate an explicit representation of the number syntax, because in both studies the syntactic effect could be explained as ad-hoc strategies based on low-level syntactic information. For example, the participants in the dictation task (Lochy et al.) could have detected syntactic markers in the verbal numbers in an early auditory stage (Fischer-Baum et al., 2018), and may have used this information to chunk the number correspondingly – either into two pairs or into a digit + triplet.

1.2.3.2. Other syntactic processes

Several studies examined particular syntactic aspects of verbal numbers that could potentially be low-level or peripheral, and thus less relevant to our research question. For example, some studies showed that to retrieve the phonological form of a number word, a phonological lexicon is accessed with two pieces of information – the digit identity and the number word lexical class (ones, tens, etc.) – the latter being syntactic information (Dotan & Friedmann, 2015; McCloskey et al., 1986). However, phonological retrieval is a language-dependent peripheral stage; clearly not the central representation that we wish to examine here. Other studies showed sensitivity to the order of number words in languages with non-canonical word order – in particular, languages such as German or Arabic, in which the ones word precedes the tens word, e.g., 123 = “one hundred

three and twenty” (Blanken et al., 1997; Brysbaert et al., 1998; Cohen et al., 1997; Contreras-Saavedra et al., 2020; Göbel et al., 2014; Hayek et al., 2020; Moeller et al., 2009; Pixner et al., 2011; Pourquié & Nespoulous, 2018; Proios et al., 2002; Qasim-Masarwa et al., 2020; Zuber et al., 2009). Nevertheless, word-order too might be handled by peripheral syntactic processes that operate on a verbal, language-specific representation rather than on a central representation (the number-reading model in Dotan & Friedmann, 2018 made precisely this assumption).

The present study focuses on verbal numbers, but syntactic mechanisms exist also for processing digit strings. Dedicated mechanisms process the order of digits – in the visual analyzer of digit strings (Dotan & Friedmann, 2018; Friedmann et al., 2010; García-Orza & Perea, 2011; Kallai & Tzelgov, 2012), as well as when writing digits (Lochy et al., 2004). Other visual analysis processes detect how many digits a number has (Dotan & Dehaene, 2020b; Dotan & Friedmann, 2018). Additional processes encode the positions of 0 and 1 (Benavides-Varela et al., 2016; Cohen & Dehaene, 1991; Dotan et al., 2021; Furumoto, 2006), two syntactically-important digits. Finally, experiments that required participants to convert multi-digit strings to quantities, e.g., using a number comparison task or a number-to-position mapping task, indicate the existence of dedicated cognitive mechanisms that implement the place-value principle of the decimal system (Dotan & Dehaene, 2020b; Nuerk & Willmes, 2005).

1.3. The present study

The present study had three goals. First, we aimed to examine whether adults maintain an explicit cognitive representation of the syntactic structure of verbal numbers, as hypothesized by McCloskey’s model. We found that they do. Second, we examined whether this syntactic representation represents the whole multi-digit number, as hypothesized in McCloskey’s model, or is limited. Our results indicate that the syntactic structure of the whole number is represented, but this representation is subject to working-memory limits. Last, we aimed to develop a methodological paradigm that can be used in the future to examine additional aspects of the syntactic structure of numbers.

The paradigm we introduce, *syntactic priming*, relies on the idea that processing a stimulus with a certain syntactic structure facilitates the processing of a subsequent stimulus having the same syntactic structure. This method was successfully used in language studies to examine the syntactic processing of sentences: the existence of a syntactic priming effect was taken to show that people represent the syntactic structure of sentences explicitly (Bock, 1986; Mahowald et al., 2016; Mehler & Carey, 1967; Pickering & Ferreira, 2008). Moreover, some studies analyzed the degree of priming between sentences with particular grammatical structures in order to reveal particular aspects of the syntactic representations of sentences (see review in Pickering & Ferreira, 2008). Nevertheless, these syntactic priming studies examined the syntactic structure of sentences, so they do not illuminate on numbers, which have a completely different syntactic structure.

We used syntactic priming to examine the syntactic structure of verbal multi-digit numbers. On each trial, participants heard a target number and responded by saying another random number. We examined the degree of similarity between the syntactic structures of the target and the response. The syntactic structure was defined as the *number word frame* – the list of lexical classes (ones, tens, etc.) in the number (Cohen & Dehaene, 1991). For example, the number word frame of “two hundred and three” is *Hundreds + Ones*. We reasoned that if the participants consistently produce responses syntactically similar to the targets, this would indicate that they explicitly represented the number's syntactic structure. Furthermore, as we shall see below, by analyzing which parts of the number's syntactic structure elicited the strongest syntactic priming effect, we could examine various aspects of this syntactic structure.

2. Method

The results of all tasks, and the code we used to analyze the data, are available in <http://osf.io/7bmqh>. The study was approved by the Tel Aviv University ethics committee.

2.1. Participants

The participants were 37 adults, aged 18-74 (mean = 33;1, SD = 12;9), with no reported cognitive disorders and with at least 12 years of education. 18 participants were native speakers of Hebrew, and the remaining 19 were native speakers of Arabic, with good knowledge of Hebrew. There was no significant age difference between the two groups ($t(30) = 1.24$, two-tailed $p = .22$; Hebrew participants: mean = 30;3, SD = 9;3; bilingual: mean = 35;10, SD = 15;0).

2.2. Stimuli and Procedure

2.2.1. Number elicitation task for syntactic priming

In each trial, the experimenter said a target number between 1 and 9,999 and the participant responded by saying a random number between 1 and 9,999, which must not be identical with the target. The task was presented to the participants as a competition game in which each party (the experimenter and the participant) says random numbers, and then they find out whether the goal was to produce the larger or the smaller number. After the participant responded, the trial's goal (larger/smaller) was revealed and the experimenter declared who "won" that trial. The experimenter recorded the winner in each trial, and at the end of each block declared the winner of the entire block, and finally of the entire experiment. Obviously, the probability to win was 50%. The experiment's real goal – to examine syntactic priming – was not disclosed to the participants.

Each participant performed 150 trials, divided into 6 blocks with 30 trials in each. There were 7, 21, 50, and 72 targets with 1, 2, 3, and 4 digits, respectively. There were 69 targets with the digit 0 and 11 targets with a teen word. All participants heard the numbers in Hebrew and responded in their native tongue – either Hebrew or Arabic. Thus, the Hebrew-speaking participants performed

the experiment in one language, and the Arabic-speaking participants performed a cross-language experiment. In the Supplementary Online Material, we describe in detail the verbal number systems in Hebrew and Arabic.

The specific list of target numbers was created with two goals in mind: to have a uniform distribution of target numbers across the valid range (1 – 9,999), and to have a uniform distribution of number word frames. Obviously, it is impossible to achieve both goals, as they contradict each other. We therefore tried to compromise between the two goals (there was no specific formal criterion). The distribution of the targets' number word frames is shown in Fig. 1.

2.2.2. Memory tasks

The participants performed a forward digit span task (Gvion & Friedmann, 2012) to tap phonological short-term memory, and a backward digit span task to tap working memory. We examined whether these measures of memory correlated with the syntactic priming effect size.

2.3. Computation of priming effects

To examine whether a participant had a significant syntactic priming effect, i.e., whether that participant tended to respond with numbers syntactically similar to the target, we used a bootstrap method. The essence of this method is simple: first, we examined the degree of syntactic similarity between the targets and the corresponding responses. Then, we examined whether the observed target-response syntactic similarity is higher than the degree of similarity expected by chance. The chance level was computed by generating a large number of random response sets and computing the target-response similarity in each of these sets.

We also examined whether the priming effect was indeed syntactic. To this end, we compared the effect of syntactic priming, as computed based on the target-response *syntactic* similarity, versus other types of priming effects, computed based on lexical, phonological, or numerical similarity. The specific algorithms we used to compute the different types of priming effects are explained below, as part of the results section.

3. Results

3.1. Data cleanup

We used 4 criteria to ensure that the participants indeed responded with random numbers and did not stick to fixed syntactic structures. For each criterion, we excluded participants with an outlier score (threshold = 75th percentile plus 150% the inter-quartile range). Overall, 5 participants were excluded. First, two participants were excluded for having a high rate of structural perseverations, i.e., trials in which the number word frame (NWF) of the response was identical with the NWF of the previous response. Second, one participant was excluded for producing too few NWFs. The specific criterion value was the average number of times each NWF was produced as response in a participant's data. Third, two participants were excluded for having a high rate of

single-word responses (i.e., round numbers). Fourth, we examined the rate of trials with response larger than the target, but no participant was excluded on these grounds. Table S3 shows the value of each of these 4 criteria for each participant.

The remaining 32 participants (16 bilingual, 16 mono-lingual) had an average structural perseveration rate of 15.2% (SD = 5.0%), an average of 9.0 responses for each NWF (SD = 1.0), and a single-word response rate of 19.6% (SD = 9.2%; Table S3 shows the per-participant statistics). Their responses were larger than the target number in 52.4% of the trials (SD = 13.1%), which was not significantly different from 50% ($t(32) = 1.03$, two-tailed $p = .31$) – i.e., they did not show any consistent bias for responding with a number larger or smaller than the target. Still, as we shall now see, the participants' responses were far from random.

3.2. A syntactic priming effect

Syntactic priming means that the participants' responses are similar to the targets they heard in terms of syntactic structure. To measure this syntactic similarity, following models of number processing (Cohen & Dehaene, 1991; Dotan & Friedmann, 2018), we defined the number's syntactic structure as the *number word frame* (NWF) – the sequence of lexical classes (ones, tens, teens, etc.) in the verbal number. In Hebrew and Arabic, in which the present study was performed, the verbal number system is slightly different from English, and so is the NWF: both in Hebrew and in Arabic, hundreds are probably not represented phonologically as a ones word plus the word “hundreds”, as is the case in English, but as single words – e.g., “five-hundred” is stored as a single phonological-lexical entry (Dotan & Friedmann, 2015)¹. For example, the Hebrew NWF of 517 is {_:hundreds} [and] {_:teens}. Similarly, each thousands word is also a single phonological-lexical entry. Namely, the number-word lexical classes in English are ones, tens, and teens; but in spoken Hebrew and Arabic, hundreds and thousands are lexical classes too.

To assess syntactic priming, we examined the syntactic structure of the responses versus that of the target numbers. Fig. 1 shows the probability for each response NWF given each target NWF. The high probabilities in the diagonal suggest that the NWF of the response tended to match that of the target number – a syntactic priming effect. We next proceeded to examine this effect statistically.

3.2.1. Syntactic priming analysis: method

3.2.1.1. General framework: bootstrap

To examine whether a participant had a significant syntactic priming effect, i.e., whether that participant tended to respond with numbers syntactically similar to the target, we used the

¹ The *written* form of the hundreds and thousands words in Hebrew and in Arabic is as two words, similar to English (five hundred = חמש מאות, /xamesh meot/). However, in spoken languages the two words are typically pronounced as one, sometimes even with a phonological modification that makes the merge easier: e.g., 3000 is written in Hebrew as שלושת אלפים, /shloshet alafim/, but is typically pronounced /shloshtalafim/.

following process. First, we examined the degree of syntactic similarity between the responses and the targets. To this end, we used a **target-response similarity function** that defined the degree of similarity between a specific target and a specific response. We used different target-response similarity functions for slightly different purposes, and they are detailed below (Section 3.2.1.2), but they all aimed to tap the similarity between the target NWF and the response NWF. A participant's overall degree of target-response syntactic similarity was defined as the average target-response syntactic similarity for that participant's 150 target-response pairs.

Fig. 1. Distribution of the responses to each target, pooled over all participants and grouped according to the target and response syntactic structure, defined as the Number Word Frame (NWF) – the sequence of lexical classes (ones, tens, etc.) in the verbal number. The NWFs of the 1- to 4-digit numbers are denoted as 1-4 characters, and specify any digit that affects the NWF (i.e., 1 in the tens positions and 0 in any position). The diagonal reflects trials in which the NWF of the response was identical with that of the target, supporting a syntactic priming effect. The histogram below the matrix shows the distribution of target NWFs, and the histogram the left shows the distribution of response NWFs.

Then, to compute whether the participant's target-response similarity index was higher than expected by chance, we used a bootstrap method. For each participant, we created a set of random responses (one per trial) using a **random response generation function**. In the simplest case, this function just generated a random response in the valid range (1 - 9,999) for each target, but we also used other random-response generation functions, as detailed below (Section 3.2.1.3). We generated 10,000 random-response sets with 150 responses in each, and for each of these random-response sets we computed the target-response syntactic similarity using the method described above. The chance-level similarity is the average target-response similarity of these 10,000 random attempts. Formally, let *SimObs* denote the average target-response similarity across the 150 real trials; and let *SimRand(i)* denote the average target-response similarity in random set #i. The priming effect size was defined, for each participant, as the *priming gain*, which reflects by how much target and response are more similar syntactically to each other in the real data than in the randomized data:

$$Gain = SimObs - average(SimRand)$$

To examine whether the syntactic priming effect was significant at the group level, we compared the per-participant priming gains to 0 using t-test. One-tailed *p* values are reported, because we assumed positive gains. To examine whether the priming effect was significant for a particular participant, we examined the bootstrap results directly:

$$p(priming) = \frac{\sum_{i=1}^{10,000} SimRand(i) > SimObs}{10,000}$$

Below, we report the average gain and the group-level significance. In Supplementary Online Material we report the per-participant priming gains and *p* values.

3.2.1.2. Target-response syntactic similarity

Syntactic priming means that the target and the response should have similar number word frames (NWFs). Thus, as an index of syntactic priming we used the degree of similarity between the target NWF and the response NWF. For each target or response, we coded whether it contained each of the 5 possible number-word lexical classes: ones, tens, teens, hundreds, and thousands. Namely, the syntactic structure of each number word was coded as 5 present/absent indicators, one per lexical class. To compute the degree of overlap between the target and the response, we considered the minimal number of modifications required in the target NWF to obtain the response NWF, given that the available modifications are: (1) Addition of a word whose lexical class was not in the target, e.g., target = 20, response = 73 (a unit word was added). (2) Omission of a lexical class, e.g., target = 73, response = 20. (3) Substitution of one lexical class with another, e.g., target=230, response=507 (the target included a tens word, the response includes a ones word instead); as long as the substituted class kept the same position in the sequence of words (e.g.,

target = 230, response = 89 cannot be considered as a substitution of a hundreds word into a ones word). The target-response dissimilarity is the number of such modifications divided by 5 (the number of possible lexical classes in our experiment), and the target-response similarity is the complement to 1. Below, we use the term “NWF match rate” to refer to this index of syntactic similarity.

For example, if the target was 305 and the response was 2470, a minimum of two modifications is required to change the target NWF (hundreds-ones) into the response NWF (thousands-hundreds-tens): adding the “thousands” class and changing “ones” to “tens”. In this example, the target-response dissimilarity is $2/5=40\%$, and the similarity is 60%. Note that this similarity measure deliberately considers only the lexical classes of the number words, and ignores the specific digit values (except 0 and 1). Note that the order of lexical classes does not cause ambiguity, because in Hebrew and Arabic numbers up to 9,999 each lexical class can appear only once.

The NWF match rate is the *degree* of syntactic similarity between the target and response. Another measure, which we used as control, examined syntactic *identity* – i.e., it considered whether the target and the response had precisely the same syntactic structure. The similarity was defined as 100% if the target and the response had precisely the same NWF, and as 0 if they did not. For example, the similarity of 305 and 350 is 0, although both contain hundreds words. The *syntactic identity rate* is the rate of trials in which the target and response had identical NWFs.

3.2.1.3. Random response generation functions

The trivial randomization method we used was *uniform random responses*: we generated, for each trial, a random response between 1 and 9,999, with equal probability for each response.

An important characteristic of this method is that it does not consider biases that the participant may have towards certain responses over others. Critically, our goal was not to discover such general biases towards particular responses, but to discover whether participants tend to produce particular numbers *as response to particular targets*. To control for this potential confound, we also used another randomization method – *random shuffling*: we shuffled the participant’s own responses in random order, i.e., we randomly paired each of the participant’s responses with one of the targets. Because this method does not change the responses but only reassigns them to targets, any overall bias the participant may have towards certain responses would be maintained also in the random-response sets, so the priming gain would not be affected by such biases.

The *uniform random responses* and the *random shuffling* methods are complementary to each other. The random shuffling method is superior in its ability to capture the participant’s tendency to match the NWF of a particular response to the particular target, while controlling for any global bias. On the other hand, the uniform random responses method is superior precisely because it can

capture such global biases, including a participant's tendency to match the NWF of their responses to the distribution of target NWFs presented in the experiment.

3.2.2. Syntactic priming analysis: results

We first analyzed the syntactic priming effect using the bootstrap method above, with syntactic similarity defined as the NWF match rate, and with uniform random responses. Across participants, the average NWF match rate was 80% (SD = 6.6%) and the average priming gain was 9.2% – significantly higher than zero ($t(31) = 7.8, p < .001$; see Table S4 for individual participant results). We also ran the bootstrap analysis with the *random shuffling* method (Section 3.2.1.3), which controls for any overall bias of the participants towards responding with particular syntactic structures. This computation too showed a significant syntactic priming effect (priming gain = 12.6%, $t(31) = 12.3, p < .001$; see Table S4 for individual participant results). The syntactic priming effect was significant also within the Hebrew-only group (with random shuffling, average NWF match rate = 80%, gain = 12.6%, $t(15) = 7.7, p < .001$) and within the Arabic-Hebrew bilingual group (average match rate = 80%, gain = 12.5%, $t(15) = 9.9, p < .001$), as well as for each individual participant except one (bootstrap $p \leq .006$ for all other participants, Holm-Bonferroni applied to correct for multiple comparisons. The per-participant gains were between 4.1% and 26.4%; Table S4). The syntactic priming effect was significant also when the target-response similarity was not defined as the NWF match rate, but instead as the NWF identity rate (Section 3.2.1.2; average gain = 25.7%, $t(31) = 8.9, p < .001$; Table S4).

Fig. 1 suggests that participants' responses tended to maintain the number length – i.e., to respond to 2-digit targets with 2-digit responses, to 3-digit targets with 3-digit responses, and so on. Could it be that the syntactic priming effect is driven by this phenomenon alone, i.e., that the priming is actually of a single aspect of the number syntax, namely the number length? We refuted this possibility by analyzing syntactic priming while controlling for the number length: we ran the bootstrap method described above, but the random shuffle method was modified to shuffle only among responses having the same length (i.e., an N-digit response was always swapped with another N-digit response). The syntactic priming effect was still significant (average gain = 3.5%, $t(31) = 6.6, p < .001$; Table S5). Similar results were obtained when we defined the response length as the number of words rather than the number of digits (average gain = 7.0%, $t(31) = 10.7, p < .001$; Table S5). We also confirmed that the priming effect did not arise from the participants representing only the existence of 0 and 1 in the number: we ran the bootstrap analysis (based on NWF match rate and the standard random shuffling), now excluding trials that had neither 0 nor 1 in the target, or neither 0 nor 1 in the response. The priming effect was significant (average gain = 11.9%, $t(31) = 10.4, p < .001$; Table S5). Thus, the participants did not represent only the number length or only the existence of 0 and 1, but several aspects of the number's syntactic structure. Below (in Section 3.4), we will see additional evidence indicating that the participants represented the presence of each of the number words – thousands, hundreds, tens, and units.

In short, the participants showed a clear and strong syntactic priming effect, which was robust across several analyses. This indicates that they represented the target number's syntactic structure (number word frame), and that this syntactic structure affected their selection of response.

3.3. The priming effect is indeed a syntactic one

We have shown a clear priming effect. To show that the origin of this priming effect was genuinely syntactic, we ruled out several alternative interpretations of the priming effect. We hereby describe each of these alternative interpretations and how it was refuted.

3.3.1. The syntactic priming effect is not an artefact of lexical priming

Lexical priming is a hypothesized tendency of the participants to occasionally repeat, within the response, one or more of the target number's words. Several arguments rule out the possibility that the syntactic priming effect was an artefact of lexical priming.

First, as we saw above, the syntactic priming effect was found also for the participants who heard numbers in Hebrew and responded in Arabic. For these participants, the response words were always different from the target words, which were in a different language, so the syntactic priming effect cannot be explained by repetition on target words.

Second, the lexical priming hypothesis assumes that syntactic priming may occur as an artefact because the participant occasionally repeats the target words, and this increases the NWF match rate. To control for this putative artefact, we recomputed the syntactic priming effect as described above (in Section 0, with NWF match rate and random shuffling) with a single change: in the computation of NWF match rate, we excluded all words that were identical in the target and the response. For example, If the target was *two hundred fifty nine* and the response was *two hundred and six*, we excluded the "two hundred". In this example, the NWF match rate, computed only for the 3 remaining words ("fifty nine" and "six"), would be $2/3 = 67\%$. As this method excludes all word repetitions, any syntactic priming effect it finds cannot be explained as an artifact of lexical priming. Still, when computed this way, the average NWF match rate over participants was 80% – significantly higher than chance (average gain = 17%, $t(31) = 12.8$, $p < .001$; Table S4).

The third argument notes that a number word is unambiguously identifiable by a lexical class (e.g., ones, teens, tens), which reflects its syntactic aspect, combined with a 1-9 digit value (McCloskey et al., 1986). For example, the combination "teens" and "5" yields the word *fifteen*. If the syntactic priming effect was actually an artifact of the participants tending to repeat a target word, or even similar words, they should repeat not only the lexical class but also the digits. Namely, the lexical priming hypothesis predicts priming of the lexical classes (which we observed as a syntactic priming effect), but critically, it also predicts a similar priming of the digit values – a "digit priming" effect. To examine this prediction we used the bootstrap process used for syntactic priming (Section 0), with random shuffling, but now the target-response similarity function was defined as the *digit match rate* – essentially, the rate of identical digits between the

target and the response. To compute the digit match rate, we first translated each verbal number, i.e., a sequence of number words, into an ordered sequence of the corresponding digits. For example, the number 517 has two words (*five* and *seventeen*) and the corresponding digit sequence is 5, 7. The target-response common digits were defined as the longest ordered sequence of digits that existed in both the target and the response (not necessarily in a consecutive manner). For example, the common digits between 1,2,4 and 2,3,4,1 are 2,4 (the digit “1” is not counted here as common because adding it would violate the relative order of digits). The digit match rate of target T and response R was defined as:

$$\text{Digit match rate} = \frac{2 \times \text{Number of common digits}}{\text{Total number of words in T and R}}$$

For example, the digit match rate of 57 and 704 is $\frac{2 \times 1}{4} = 50\%$ (one common digit: 7), and the digit match rate of 124 and 2341 is $\frac{2 \times 2}{7} = 57\%$ (two common digits: 2,4).

The digit priming effect was significant (average digit match rate = 23.9%, gain = 5%, $t(31) = 3.76$, $p < .001$; for the Hebrew-only group, gain = 4.3%, $t(15) = 2.23$, $p = .04$; for the bilingual group, gain = 5.6%, $t(15) = 3.04$, $p = .01$; Table S6). Nevertheless, the digit priming effect was significantly weaker than the syntactic priming effect: the per-participant p values of the syntactic priming were lower than the p values of the digit priming (paired $t(31) = 4.04$, one-tailed $p < .001$; Hebrew-only group, paired $t(15) = 3.31$, $p = .002$; Bilingual group, paired $t(15) = 2.34$, $p = .02$; we compared the p values rather than the gains because the similarity functions in the two analyses are different, so the gains are not comparable directly). At the individual participant level too, the digit priming effect was significant only for 13 of the 32 participants (bootstrap $p < .05$ after Holm-Bonferroni correction).

Altogether, these results show that the syntactic priming effect is not an artefact of lexical priming.

3.3.2. The syntactic priming effect is not an artefact of phonological priming

Phonological priming is a hypothesized tendency to respond with words that are phonologically similar to the target words. Several arguments indicate that the syntactic priming effect cannot be explained as an artefact of phonological priming.

The first argument initially notes that in Hebrew, each number word is created by merging a stem, denoting the digit, with a suffix denoting the word’s lexical class (e.g., /im/ for tens: 3 = /shalosh/, 30 = /shloshim/). Syntactic priming essentially means that the participants tend to repeat the target words’ suffixes. Phonological priming predicts that the responses will phonologically resemble not only the target’s suffixes but also the stems of the target words.

To assess this prediction, we used our bootstrap method (Section 0) with random shuffling, but now the target-response similarity function indexed phonological similarity. To focus on phonological (rather than syntactic) similarity, we examined the similarity between the stems of

the number words in the target and response, and we ignored the affixes (as an example in English, for the word “fifty” we would ignore the suffix “-ty” and consider only the stem “fif”). The target-response *phoneme match rate* was computed as follows. Let T and R denote the stems of the number words in the target and in the response, correspondingly. Then:

$$\text{Phoneme match rate} = \frac{2 \times \text{No. of phonemes that appeared in both T and R}}{\text{Total number of phonemes in T and R}}$$

To compute the number of phonemes that appeared in both T and R, we counted pairs of identical target-response phonemes, subject to the limit that each phoneme can only be counted in one pair. For example, the numbers *four* (phonemes: /f/, /o/, /r/) and *fifty* (/f/, /i/, /f/, without the suffix /ti/) have a total of 6 phonemes, and only one pair of shared phonemes (/f/). Of the two /f/ phonemes in “fifty”, only one can be paired with “four”, because “four” only has one /f/. The phoneme match rate of “four” and “fifty” is therefore $\frac{2 \times 1}{6} = 33\%$.

Phonological priming was not computed for the bilingual group, because the phonology of Hebrew and Arabic is somewhat different, and may be perceived in different ways for different participants, depending on their accent and on their fluency in Hebrew.

If the priming effect originates in phonological similarity, the phonological priming effect should be as strong as the syntactic priming effect (Section 3.2.2). A phonological priming effect was observed (average gain = 4.2%, $t(15) = 3.5$, $p = .003$; Table S7), however, the effect was relatively weak, and significantly weaker than the syntactic priming effect (the phonological priming p values were higher than the syntactic priming p values; paired $t(15) = 2.54$, one-tailed $p = .01$). Moreover, at the individual participant level, only half of the participants showed significant phonological priming after Holm-Bonferroni correction). This weak phonological priming effect could be completely reduced to digit priming, i.e., repetition on specific digits: when controlling for digit-repetition priming by excluding digits that were identical in the target and the response (same method as used above in Section 3.3.1), no group-level phonological priming effect was observed any longer (gain < 0), and only one participant showed a significant priming effect (Table S7). In short, the stems of the target and response words did not show reliable phonological similarity.

Second, if the syntactic priming effect is an artefact of phonological similarity, the bilingual group should show less priming for the teens word than for the tens words. The reason is that the tens numbers are more similar between Hebrew and Arabic than the teens are. More specifically, in both languages the tens and the teens are marked by specific suffixes; the tens suffixes are phonologically similar across the two languages, but the teens suffixes are phonologically different. For example, 50 is /xamifim/ in Hebrew and /xamsin/ in Arabic, whereas 15 is /xamef-esre/ in Hebrew and /xamistaf/ in Arabic. In short, if the syntactic priming effect originates in phonology, in the bilingual group it should be stronger for the tens than for the teens. To assess

this prediction, we computed the degree of syntactic priming separately for the tens and for the teens. For each of the two lexical classes, we used our bootstrap method with random shuffle (Section 0). We analyzed only items in which the target included the lexical class in question (either tens or teens). When creating the random-response sets, we first generated random responses for all 150 trials and only then applied this filtering. As target-response similarity function, we defined similarity as 100% if the response too included the lexical class in question, and as 0 if it did not. Contrary to the prediction of the phonological priming hypothesis, the teens priming effect in the bilingual group (average gain = 21.5%) was actually stronger than the tens priming effect (average gain = 5.7%; paired $t(15) = 4.7$, two-tailed $p < .001$; Table S8).

Third, if the responses are dominated by phonological priming, both syntactic priming and lexical priming effects should be higher in the Hebrew-only group than in the bilingual group, because each Hebrew number word is more similar phonologically to itself than to the corresponding word in Arabic. Contrary to this prediction, the syntactic priming effect was similar in the two groups (in both groups, 15 out of the 16 participants had bootstrap $p = 0$; comparing the per-participant gains, unpaired $t(30) = 0.05$, two-tailed $p = 0.95$). The lexical priming effect, measured via the digit match rate (Section 3.3.1), was similar across the two groups too (numerically it was even stronger in the bilingual group, with an average gain of 5.3% versus 4.3% in the Hebrew group, but this difference was not significant: comparing the per-participant gains between the two groups, unpaired $t(30) = 0.51$, two-tailed $p = .62$).

Altogether, these results show that the syntactic priming effect is not an artefact of phonological priming.

3.3.3. The syntactic priming effect is not an artefact of numerical distance

Another alternative hypothesis attributes the syntactic priming to a putative tendency to respond with numbers that are numerically close to the target (an anchoring effect, Tversky & Kahneman, 1974). Close numbers tend to have similar number word frames (e.g., they are likely to have the same thousand and hundred digits), so a strategy of choosing a close-to-target response could give rise to a syntactic priming effect in our analysis.

However, such distance effect cannot account for the syntactic priming effect, because the syntactic priming effect survived the inclusion of numerical distance as a controlled factor. To show this, we recomputed the syntactic priming effect (as in Section 0, with similarity defined as the NWF match rate), but crucially, we used a random-response generation function that ensures that the target-response numerical distance in the randomized sets would be similar to the participant's real target-response numerical distance. We ran two different bootstrap analyses, with two different random-response generation functions.

The first randomization function we used relied on the participant-specific distribution of target-response distances. For each participant, we computed the mean and the standard deviation of the target-response distances based on the participant's real responses. Then, we chose random

responses according to a normal distribution function having the same mean and standard deviation as in the participant’s real data. The syntactic priming effect computed this way was significant (average gain = 6.1%, $t(31) = 5.95$, $p < .001$, Table S9), indicating that the syntactic priming was not an artefact of numerical distance.

The second random-response generation function shuffled each participant’s actual responses, similar to the random shuffle method described in Section 3.2.1.3. By using random shuffle instead of random responses, we controlled for any possible target-independent bias towards particular responses. Importantly, this method maintained the mean and standard deviation of the participant’s target-response distances. The mean distance is unchanged by shuffling, and the specific shuffling algorithm we used was designed to ensure that the standard deviation of distances in the shuffled data would be as close as possible to the real data.

The shuffling algorithm consisted of 3 stages. The first stage shuffled the participant’s responses randomly (as in Section 3.2.1.3). Two additional stages aimed to obtain target-response distance distribution similar to the real data. The first improvement stage was a greedy algorithm aiming to maximize the number of randomized trials in which the target-response distance was close to a corresponding target-response distance in the real data. The specific algorithm was as follows. We processed the participant’s shuffled trials one at a time, in random order. If the target-response distance of the shuffled trial was close enough to one of the target-response distance in the real data ($|distance_{shuffled} - distance_{real}| \leq 10$), the shuffled trial was considered as valid. In this “distance-matching”, each real-data distance could be “used” only once – e.g., taking an extreme case as an example, it is invalid to create 150 target-response distances that are similar to each other and to one of the real target-response distances, while ignoring the 149 other real target-response distances. If the target-response distance of a particular shuffled trial was invalid (i.e., it could not be matched with any target-response pair in the real data such that $|\Delta distances| \leq 10$), the algorithm tried to swap the response of that trial with another random trial, such that both would have valid distances. This matching-and-swapping procedure was done repeatedly for all trials (in random order) until no more “improving swaps” were possible.

The second improvement stage used a *simulated annealing* procedure (*simeanneal* Python module, Perry, 2019). Simulated annealing is an optimization algorithm that runs in several iterations; in each iteration it attempts a random change and accepts the change in a probability that depends on the change’s effect on a cost function. In our case, the change was swapping two random trials, and the cost function was the difference between the standard deviation of the 150 target-response distances in the shuffled trials and the standard deviation in the real data. The effect on $\Delta cost$ on the likelihood of accepting the change is probabilistic: the more the change reduces the cost, the more likely it is to be accepted, but cost-increasing changes can also be accepted occasionally. The probability to accept a change also depends on time: as the process continues

and we presumably get closer to the optimal solution, the algorithm decreases the probability to accept cost-increasing changes.

We verified that this shuffling method indeed maintained the mean and standard deviations of the target-response distances. The mean was obviously identical with the real data, as it is not affected by shuffling; and the differences in standard deviations between the real and shuffled data were small: excluding two participants, the difference did not exceed 5% in any bootstrap iteration; and for 27 participants it never exceeded 3% (Table S10).

This analysis showed a significant syntactic priming effect (average gain = 8.1%, $t(31) = 11.7$, $p < .001$, Table S9). Thus, with both methods we used to control for numerical distance, the results were clear: the syntactic priming effect cannot be explained as an artefact of the target-response numerical distance.

Fig. 2. The distribution of response-target distances (Gaussian smoothing with $\sigma = 10$ numerical units was applied). Positive values indicate response larger than target (over all participants, this was the case in 53% of the trials). The distance (anchoring) effect is clearly visible, however, it does not explain the syntactic priming.

Although the numerical distance could not account for the syntactic priming effect, it did affect the participant's selection of responses: the participants tended to pick responses that were numerically close to the target (Fig. 2; this effect becomes even clearer when plotting the likelihood of each target-response pair, Fig. 3). To show that the distance effect was statistically significant, we used our bootstrap method, with the dependent variable indexing numerical distance: the target-response similarity function was defined as $-\log(|\text{target} - \text{response}|)$ (and as 0 if the target and response were equal). Random responses were generated using the random shuffling method (Section 3.2.1.3), with an additional limitation – that the shuffling maintains the real response's NWF in each trial. Maintaining the NWF ensures that any observed distance effect would not be an artefact of syntactic priming. The distance effect computed this way was significant (average gain in terms of $\log(\text{distance}) = 0.76$, $t(31) = 6.44$, $p < .001$; Table S9). Namely, the participants' selection of responses was affected both by syntactic priming and by the numerical distance.

Fig. 3. The distribution of targets and responses, pooled over participants. The data was smoothed by ignoring the unit digit, thereby summing each 10x10 data points into a single matrix pixel. Each point in this graph shows the conditional probability for a particular response given a particular target, i.e., each column in the matrix sums to 100%. Below and near the matrix: the distributions of the targets irrespectively of the response, and of the responses irrespectively of the target, presented as the probability per numerical unit (Gaussian smoothing with $\sigma = 10$ numerical units was applied to these histograms).

3.3.4. Interim summary

The analyses above clearly show that the syntactic priming effect could not be explained as an artefact of lexical priming, phonological priming, or numerical distance. Thus, the effect is genuinely syntactic, and indicates that the participants represented the number's syntactic structure. We next examine the characteristics of this syntactic representation.

3.4. Number syntax and working memory

We examined whether the syntactic priming effect was elicited by particular syntactic elements – i.e., by words in particular positions in the number. We started by computing the priming effect separately for each lexical class – ones, tens, hundreds, and thousands. We used the method described in Section 3.3.2: for each lexical class, we ran a separate bootstrap analysis with random shuffle, while including only trials in which the target included the lexical class in question. Target-response similarity was defined as 100% if the response too included the lexical class in question, and as 0 if it did not. As Table 1 shows, the syntactic priming effect (priming gain) was the strongest for the thousands words and continuously decreased, with the weakest effect for the ones words.

Table 1. The syntactic priming effect (priming gain) for each lexical class. The priming effect is the strongest for the thousands, and decreases monotonously for “lower” classes.

	Thousands	Hundreds	Teens	Tens	Ones
NWF Match rate	77%	79%	28%	70%	64%
Gain	35%	14%	19%	7%	5%
t(31)*	15.2	13.4	6.0	6.4	5.0

* Comparing the gains versus 0 with t-test. All $p < .001$.

This dependency of the syntactic priming effect size on the lexical class can be interpreted in two ways. One interpretation is that the “larger” lexical classes (e.g., tens as opposed to ones) elicited more priming because their cognitive representation is different. For example, they may be more salient because their impact on the quantity is larger. Another interpretation is that the priming effect size does not depend on the lexical class per-se, but on the position of the word in the verbal number, with stronger priming for earlier words. Such serial-position pattern could arise, for example, from the involvement of working memory mechanisms. The idea is that representing the verbal number’s syntactic structure may require working memory resources (Bahnmüller, Galefski, Moeller, Nuerk, & Artemenko, in prep.), and a working memory limit may disrupt the later words more than the earlier ones (a primacy effect, Atkinson & Shiffrin, 1968). As a result, later words would be under-represented in the syntactic structure, and would therefore elicit a weaker syntactic priming effect.

To arbitrate between the lexical-class interpretation and the serial-position interpretation, we examined whether the degree of syntactic priming is best predicted by the lexical class or by the word’s position in the verbal number. These two parameters sometimes dissociate due to the presence of 0 (e.g., the word “twenty” has the same lexical class but different positions in 4,021 and 4,321, and so does the word “one”). Fig. 4 clearly indicates that the degree syntactic priming was modulated by the word position in the verbal number, whereas the effect of lexical class was not monotonous.

Fig. 4. The match rate for words in particular positions, according to their lexical class and position in the verbal number. The match rate decreases monotonously for words in later positions in the number, but the effect of lexical class is not monotonous.

To examine statistically whether the syntactic priming effect of particular number words was affected by their lexical class or by their position, we used the following analysis. We defined the priming effect for each number word in the target as the *syntactic match*: 1 if the participant’s response included a word having the same class as the target word, and 0 if it did not. Essentially, the syntactic match denotes whether the particular target word contributed to syntactic priming or not. We then submitted the syntactic match data to logistic linear mixed model (LLMM) with the participant as a random factor and with three within-subject factors: the word’s position in the target number as a numeric factor (1 = first), the word’s lexical class as a numeric factor (1 = thousand, 2 = hundreds, 3 = tens, 4 = ones), and the number length (how many digits it has) as a non-numeric factor. We used R (R Core Team, 2019) with the *lme4* package (Bates et al., 2015). To determine whether the effect of a particular factor was significant, we used a log likelihood ratio test that compared the LLMM to a model that was identical but without the factor in question. For these comparisons, we report the test statistic $2(LL_0 - LL_1)$, which follows a χ^2 distribution (LL_0 and LL_1 denote the log-likelihoods of the reduced model and the full model), the corresponding p value, and the odds ratio (the exponent of the factor’s coefficient in the full model). The Position factor had a significant negative effect (Table 2) – i.e., the early words in each target elicited strong syntactic priming, and later words elicited smaller priming effects. In contrast, the effect of Lexical Class was positive (odds ratio > 1), i.e., in the model that included both factors, the likelihood for a syntactic match was not decreased by the word having a “later” lexical class (the “Ones” class being the latest), which would explain the effect observed in Table

2, but was actually increased. Namely, the degree of priming was affected by the word's serial position in the verbal number, not by its lexical class.

Table 2. The effect of the target word position and its lexical class on syntactic priming in the linear mixed model (dependent variable = per-word syntactic match). The position effect was significant, but the lexical class effect was opposite to the predicted direction.

Predictor	Odds ratio	95% confidence interval	Significance
(Intercept)	2.18	1.53 - 3.11	
Position	0.53	0.47 – 0.60	$\chi^2(1) = 111.8, p < .001$
Lexical class (numeric)	1.19	1.08 – 1.31	$\chi^2(1) = 12.4, p < .001$
Number length			$\chi^2(3) = 66.2, p < .001$
Length = 2	2.11	1.68 – 2.65	
Length = 3	2.72	2.08 – 3.58	
Length = 4	3.81	2.75 – 5.28	
<i>Random effects</i>			
σ^2 within participants	3.29		
σ^2 between participants	0.36		

The finding that serial position, not lexical class, is responsible for the syntactic priming slope in Table 1, agrees with the idea that the syntactic priming (and the syntactic representation responsible for it) is modulated by short-term memory (STM) or working memory (WM): the participants seem to remember the earlier parts of the syntactic structure better than the later parts. An additional question concerns the type of these short-term memory processes, which are needed for syntactic representation: are they domain-general short-term memory processes, or a memory process dedicated to representing number syntax? If the memory processes in question are domain-general, a participant's sensitivity to the word position should be higher for participants with lower STM or WM. However, this was not the case. We defined the participant's sensitivity to word position as the slope of the priming-gain-by-position curve, computed by regressing, for each participant, the priming gain against serial position. We then examined whether participants with lower memory spans had steeper gain-by-position slopes (i.e., positive correlation, as the slope was negative). We used multiple linear regression to regress the gain-by-position slope against the participants' forward digit span (an index of short-term memory), their backward digit span (an index of working memory), and the language group (bilingual or monolingual). This analysis was done only for 17 participants, because the remaining participants were no longer available at this time of the study. Neither span predictor had a significant effect (forward digit span, $b = -.002$, i.e., opposite to the predicted direction; backward digit span, $b = .001$, one-tailed $p = .46$). These results may suggest that the memory-sensitive syntactic process is not a domain-general memory process. However, admittedly, to reach a reliable conclusion regarding the precise nature of the

relation between syntactic representation and short-term/working memory, this issue should be examined more deeply and systematically than the analysis we could do with our present data.

4. Discussion

4.1. An explicit representation of the number's syntax

The first goal of this study was to examine whether the cognitive system represents explicitly the syntactic structure of verbal numbers. The participants heard a target number in each trial and responded with a random number. They clearly represented the number's syntactic structure, because the syntactic structure of their response resembled that of the target. In particular, the participants often responded with numbers having the same lexical classes (ones, tens, teens, etc.) as the classes that appeared in the target. This priming effect clearly shows that the participants represented the lexical classes, i.e., the syntactic structure, of both the target and the response, and could transfer this syntactic information from the verbal input processes (the target) to the verbal production stage (the response).

Importantly, we refuted several alternative non-syntactic interpretations of this priming effect. First, we showed that the syntactic priming was not an artefact of lexical priming – a tendency to include some of the target number's words in the response. Contrary to the idea of lexical priming, (a) the syntactic priming effect was observed even when the target and the response were in different languages, i.e., when they were always different words; (b) the syntactic priming effect was observed even when we excluded words that were identical in the target and the response, thereby eliminating any possible effect of word repetition and lexical priming; and (c) the syntactic priming, i.e., the priming of the number's lexical classes, was stronger than the priming of digit values, whereas lexical priming predicts priming of both.

Second, we showed that the syntactic priming was not an artefact of phonological priming – a tendency to respond to a target number with similar-sounding words. Contrary to the idea of phonological priming, (a) the similarity of response words to target words was observed in the words' morphological affixes, which reflect the number's syntax, but there was only a small and non-reliable phonological priming in the number word stems; (b) the bilingual group showed stronger priming effect for teens than for tens, whereas the phonology of these words predicts the opposite, because the Hebrew-Arabic phonological similarity is larger for tens than for teens; and (c) the priming effect was not stronger for monolingual participants than for bilingual participants, as should have been the case had it been driven by phonology.

Third, the syntactic priming was not an artefact of an anchoring effect (Tversky & Kahneman, 1974) – a tendency to respond with numbers numerically close to the target. Contrary to the idea of distance-based priming, the syntactic priming effect was observed even when we controlled for the target-response numerical distance. Nevertheless, we did observe an effect of numerical

distance on top of the syntactic priming effect. Namely, the participants' responses were affected both by syntactic priming and by numerical distance as two separate factors.

Having refuted all these alternative interpretations, we could reliably conclude that the priming effect was a syntactic one.

Several neuropsychological studies showed that the processing of the number's syntax can be selectively impaired or selectively spared (Blanken et al., 1997; Cipolotti, 1995; Cipolotti et al., 1994; Dotan et al., 2014; Furumoto, 2006; Noël & Seron, 1993). These studies indicate the existence of dedicated mechanisms that process the syntax of numbers. Additional studies showed that syntactic processing exists not only in general, but even specifically within the verbal-number processing mechanisms (Dotan & Friedmann, 2018; Lochy et al., 2004). The present study joins these studies in showing that the human cognitive system processes the syntactic structure of verbal numbers. Importantly, the present study also extends these previous studies. First, our data shows that the syntax of verbal numbers is not merely "handled" in some way, e.g., in processes that handle low-level syntactic information, perhaps even merely for single number words, but is represented explicitly – as a series of number-word lexical classes or in an isomorphic representation. Second, our data suggests that the syntactic representation is not comprehension-specific or production-specific, but a central one. In our paradigm, the participants heard a number in each trial, and its syntactic structure primed the spoken number's syntactic structure. Thus, the syntactic representation is not an input-only or output-only representation, but exists for both. The most likely interpretation of the comprehension-to-production priming is a central syntactic representation, accessed by both the input and the output processes, as McCloskey (1992) proposed. Third, our data suggests that the syntactic representation is capped by working memory, as detailed below.

4.2. The syntactic representation is capped by memory

Our second goal was to examine whether the syntactic representation incorporates the whole number or just parts of it. We showed that at least for 4-digit numbers, the full number is represented: our participants showed a significant priming effect for all words, including the ones word. Nevertheless, our findings also suggest that the syntactic representation is limited by working memory: the priming effect was stronger for words occurring earlier in the number, and weaker for later words. Crucially, we showed that this position effect could not be explained as stronger priming for particular lexical classes, but was driven by the word's serial position in the verbal number, in line with the idea of a working-memory-capped representation. We acknowledge that our experiment included numbers only up to 4 digits long; it may be that with longer numbers, some words would have failed to show a syntactic priming effect.

The decline of syntactic priming with word position suggests a memory-limited representation, yet this position effect did not correlate with the participants' short-term memory or working

memory measures. This perhaps suggests that representing the number's syntax involves a dedicated memory process, which is particular to handling number syntax and is not involved in general memory tasks such as the tasks that we used here (forward/backward digit span). In language too, the working memory needed to represent the syntax of sentences seems to be a dedicated mechanism, dissociable from phonological working memory (Friedmann & Gvion, 2003). Future studies may look more deeply into the putative role of working memory in handling the number's syntax.

The idea that representing the number syntax requires working memory resources is in good agreement with a study that suggested the involvement of working memory in syntactic processing during number dictation (Zuber et al., 2009). In some languages, e.g., German, digit strings are written as in English, but in the verbal number the ones word precedes the tens word (i.e., you read 234 as “two hundred, four and thirty”). Zuber et al. showed that in a number dictation task in German, children sometimes transposed the decade and the unit digits, presumably because they wrote the digits from left to right in the same order in which they heard the words. Critically, the probability for such inversion errors correlated with the children's performance in a working memory task. Thus, it seems that the syntactic operation of inverting the word order, which is necessary to write numbers correctly in German, requires working memory resources. Nevertheless, as we noted in the Introduction, even if number-word inversion (Zuber et al.) and a central representation (the present findings) can both be referred to as “number syntax”, these are two different operations which may invoke different cognitive processes (Dotan & Friedmann, 2018).

4.3. Serial or hierarchical syntactic representation?

The present study defined a verbal number's syntactic structure as the number word frame – the series of lexical classes in the number. Some models of number processing, including McCloskey's, assume that the deep representation of the number's syntax is not serial, as we described here, but hierarchical (Barash & Dotan, 2019; Dotan & Friedmann, 2018; McCloskey, 1992). Importantly, such hierarchical models of number syntax do not contradict our conclusions but complement them, for two reasons. First, we showed that the participants represented the number's syntactic structure, but our findings can be easily interpreted also under a hierarchical-syntax model. We described the number syntax using serial rather than hierarchical terminology merely because the serial description is simpler. Second, hierarchical and serial representations of syntax may co-exist side by side in separate processing stages. Indeed, some models of number processing assume precisely this – that the number's syntactic structure is represented in a hierarchical manner, and this hierarchical representation is then translated into a serial representation – the number word frame (Cohen & Dehaene, 1991; Dotan & Friedmann, 2018).

A related question is whether the number word frame, on which our analyses focused, should be described as a syntactic representation or a morphological one. On the one hand, the NWF

specifies the lexical class of each word, and decomposing of a number word into a class and a digit value (or constructing a word from these two parameters) is best described as a morphological operation. On the other hand, the NWF reflects the relation between the words in a multi-digit number, and in this sense it represents a syntactic structure. It could also be possible that the structure of multi-digit numbers is handled by multiple processes, some of which are better described as syntactic and others as morphological. Future studies may ask whether the syntax and morphology of verbal numbers are indeed separate cognitive constructs, or only two different terminologies that describe the same cognitive processes.

4.4. Syntactic priming as a method to investigate the syntactic representation of numbers

The last goal of this study was to develop a paradigm that would allow examining number syntax in detail. We propose that the paradigm we introduced here, syntactic priming, can do precisely that; and that it could become a powerful method to look into additional aspects of the syntactic representation of numbers.

Previous studies used different methods to examine the syntactic processing of verbal numbers – e.g., neuropsychological examination of individuals with selective deficits (Dotan & Friedmann, 2018; Marangolo et al., 2004; McCloskey et al., 1986), manipulating the grammaticality of number-word sequences (Brutman & Dotan, 2019; Hung et al., 2015), saying the same numbers with different verbal structure (Lochy et al., 2002; Noël & Seron, 1997), and examining languages with tens-ones number-word inversion (Blanken et al., 1997; Brysbaert et al., 1998; Cohen et al., 1997; Contreras-Saavedra et al., 2020; Göbel et al., 2014; Hayek et al., 2020; Moeller et al., 2009; Pixner et al., 2011; Pourquié & Nespoulous, 2018; Proios et al., 2002; Qasim-Masarwa et al., 2020; Zuber et al., 2009). Syntactic priming is an additional paradigm, which offers unique advantages: first, it does not merely tap “sensitivity” to the number syntax, a phenomenon that can occur also for low-level syntactic processes, perhaps even such that handle syntactic information merely for single digits or single number words. Rather, syntactic priming can tap the full syntax of a multi-digit number. Second, syntactic priming can be used to reveal several specific properties of the syntactic representation. Here we examined one such property – the representation being limited by working memory. Other variants of syntactic priming may examine additional properties of number syntax – e.g., whether the syntactic representation is abstract and/or hierarchical (as hypothesized by McCloskey et al., 1986), whether and how the syntactic processing of numbers is related with the syntactic processing of sentences, and so on. Moreover, because the method is simple and does not require much training, it may also be used with young children to examine the development of number syntax.

In the present study, the participants could say any number that came to their mind, and were instructed only “to choose a random number”. This free-response design has several advantages – e.g., it showed that the syntactic representation serves comprehension as well as production

processes, and it allowed examining various properties of the number syntax rather than focusing on just one aspect. At the same time, the free-response design also has limitations. A major limitation of this design is that unlike classical priming experiments, in which the experimenter typically manipulates both the prime and the target, here we could manipulate the prime but not the stimulus that follows it (which was chosen freely by the participants). Future syntactic priming studies may use more structured designs, e.g., by choosing both the prime and the target, or by imposing various restrictions on the participant's responses, in order to examine specific aspects of the number syntax.

Another methodological issue concerns the selection of specific targets. This issue is important because the participants may be affected not only by the target of the specific trial, but also by the distribution of number word frames in the experiment. For example, this may be one of the reasons for the high rate of responses without zeros (see Fig. 1) – such numbers were also more frequent as targets. Indeed, other studies too showed that participants are sensitive to the distribution of stimuli (Cicchini et al., 2014; Dotan & Dehaene, 2020a; Hanks et al., 2011; Kording & Wolpert, 2004; Summerfield & de Lange, 2014). The question of target distribution was not in the scope of the present study; to select the specific list of targets, we used a vague compromise between uniform distribution of numbers and uniform distribution of number word frames. Future studies may examine more systematically how syntactic priming is affected by a specific trial versus the experimental context; for example, by using different distributions of targets.

4.5. Conclusion

Processing syntax, in particular that of numbers, is a big cognitive challenge, and correspondingly several different cognitive mechanisms are involved in this task. The big question is how precisely the human cognitive system handles this challenge, i.e., how do these syntactic mechanisms operate. To answer this question, we should look in detail into specific syntactic processes and the details of specific syntactic representations. The present study offers an initial step in this direction: we showed an explicit central representation of the number syntax, which is capped by memory limits. We hope that future studies can use the method we introduced here to examine this representation in more detail.

References

- Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *Psychology of learning and motivation: Advances in research and theory*, vol. 2 (pp. 89–195). Academic Press.
- Bahnmueller, J., Galefski, S., Moeller, K., Nuerk, H.-C., & Artemenko, C. (n.d.). Re-inverting inversion: Natural offloading in number transcoding? *Manuscript in Preparation*.
- Barash, T., & Dotan, D. (2019). *Writing multi-digit numbers is a structural-hierarchical process*.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>

- Benavides-Varela, S., Passarini, L., Butterworth, B., Rolma, G., Burgio, F., Pitteri, M., Meneghello, F., Shallice, T., & Semenza, C. (2016). Zero in the brain: A voxel-based lesion-symptom mapping study in right hemisphere damaged patients. *Cortex*, 77, 38–53. <https://doi.org/10.1016/j.cortex.2016.01.011>
- Blanken, G., Dorn, M., & Sinn, H. (1997). Inversion errors in Arabic number reading: Is there a nonsemantic route? *Brain and Cognition*, 34(3), 404–423. <https://doi.org/10.1006/brcg.1997.0917>
- Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology*, 18(3), 355–387. [https://doi.org/https://doi.org/10.1016/0010-0285\(86\)90004-6](https://doi.org/https://doi.org/10.1016/0010-0285(86)90004-6)
- Brutman, N., & Dotan, D. (2019). *Structural chunking as evidence for explicit representation of the number's verbal structure*.
- 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. [https://doi.org/10.1016/S0010-0277\(98\)00006-7](https://doi.org/10.1016/S0010-0277(98)00006-7)
- Campbell, J. I. D., & Clark, J. M. (1992). Cognitive number processing: An encoding-complex perspective. In J. I. D. B. T.-A. in P. Campbell (Ed.), *The Nature and Origins of Mathematical Skills* (Vol. 91, pp. 457–491). North-Holland. [https://doi.org/https://doi.org/10.1016/S0166-4115\(08\)60894-8](https://doi.org/https://doi.org/10.1016/S0166-4115(08)60894-8)
- Cappelletti, M., Kopelman, M. D., Morton, J., & Butterworth, B. (2005). Dissociations in numerical abilities revealed by progressive cognitive decline in a patient with semantic dementia. *Cognitive Neuropsychology*, 22(7), 771–793. <https://doi.org/10.1080/02643290442000293>
- Cheung, P., & Ansari, D. (2020). Cracking the code of place value: The relationship between place and value takes years to master. *PsyArXiv Preprints*. <https://doi.org/10.31234/osf.io/zqpdw>
- Cicchini, G. M., Anobile, G., & Burr, D. C. (2014). Compressive mapping of number to space reflects dynamic encoding mechanisms, not static logarithmic transform. *Proceedings of the National Academy of Sciences of the United States of America*, 111(21), 7867–7872. <https://doi.org/10.1073/pnas.1402785111>
- Cipolotti, L. (1995). Multiple routes for reading words, why not numbers? Evidence from a case of Arabic numeral dyslexia. *Cognitive Neuropsychology*, 12(3), 313–342. <https://doi.org/10.1080/02643299508252001>
- Cipolotti, L., Butterworth, B., & Warrington, E. K. (1994). From “one thousand nine hundred and forty-five” to 1000,945. *Neuropsychologia*, 32(4), 503–509. [https://doi.org/10.1016/0028-3932\(94\)90094-9](https://doi.org/10.1016/0028-3932(94)90094-9)
- Cohen, L., & Dehaene, S. (1991). Neglect dyslexia for numbers? A case report. *Cognitive Neuropsychology*, 8(1), 39–58. <https://doi.org/10.1080/02643299108253366>
- Cohen, L., & Dehaene, S. (2000). Calculating without reading: Unsuspected residual abilities in pure alexia. *Cognitive Neuropsychology*, 17(6), 563–583. <https://doi.org/10.1080/02643290050110656>
- Cohen, L., Verstichel, P., & Dehaene, S. (1997). Neologistic jargon sparing numbers: A category-specific phonological impairment. *Cognitive Neuropsychology*, 14(7), 1029–1061. <https://doi.org/10.1080/026432997381349>
- Contreras-Saavedra, C. E., Willmes, K., Koch, I., Schuch, S., Benini, E., & Philipp, A. M. (2020). Multilingual two-digit number naming: The influence of composition rules on language switching. *Quarterly Journal of Experimental Psychology*, 73(9), 1481–1494. <https://doi.org/10.1177/1747021820916108>
- Dehaene, S. (1992). Varieties of numerical abilities. *Cognition*, 44(1–2), 1–42. [https://doi.org/10.1016/0010-0277\(92\)90049-N](https://doi.org/10.1016/0010-0277(92)90049-N)
- Dehaene, S., & Cohen, L. (1995). Towards an anatomical and functional model of number processing. *Mathematical Cognition*, 1, 83–120.
- Dehaene, S., Meyniel, F., W, C., Wang, L., & Pallier, C. (2015). The neural representation of sequences: From transition probabilities to algebraic patterns and linguistic trees. *Neuron*, 88(1), 2–19. <https://doi.org/10.1016/j.neuron.2015.09.019>
- Delazer, M., & Bartha, L. (2001). Transcoding and calculation in aphasia. *Aphasiology*, 15(7), 649–679. <https://doi.org/10.1080/02687040143000104>
- Dotan, D., & Dehaene, S. (2016). On the origins of logarithmic number-to-position mapping. *Psychological Review*, 123(6), 637–666. <https://doi.org/10.1037/rev0000038>
- Dotan, D., & Dehaene, S. (2020a). Tracking priors and their replacement: Mental dynamics of decision making in the number-line task. *PsyArXiv*. <https://doi.org/10.31234/osf.io/skcg4>

- Dotan, D., & Dehaene, S. (2020b). Parallel and serial processes in number-to-quantity conversion. *Cognition*, 204, 104387. <https://doi.org/10.1016/j.cognition.2020.104387>
- Dotan, D., Eliahou, O., & Cohen, S. (2021). Serial and syntactic processing in the visual analysis of multi-digit numbers. *Cortex*, 134, 162–180. <https://doi.org/10.1016/j.cortex.2020.10.012>
- Dotan, D., & Friedmann, N. (2015). Steps towards understanding the phonological output buffer and its role in the production of numbers, morphemes, and function words. *Cortex*, 63, 317–351. <https://doi.org/10.1016/j.cortex.2014.08.014>
- Dotan, D., & Friedmann, N. (2018). A cognitive model for multidigit number reading: Inferences from individuals with selective impairments. *Cortex*, 101, 249–281. <https://doi.org/10.1016/j.cortex.2017.10.025>
- Dotan, D., Friedmann, N., & Dehaene, S. (2014). Breaking down number syntax: Spared comprehension of multi-digit numbers in a patient with impaired digit-to-word conversion. *Cortex*, 59, 62–73. <https://doi.org/10.1016/j.cortex.2014.07.005>
- Fischer-Baum, S., Mis, R., & Dial, H. (2018). Word deafness with preserved number word perception. *Cognitive Neuropsychology*, 35(8), 415–429. <https://doi.org/10.1080/02643294.2018.1515734>
- Friedmann, N., Dotan, D., & Rahamim, E. (2010). Is the visual analyzer orthographic-specific? Reading words and numbers in letter position dyslexia. *Cortex*, 46(8), 982–1004. <https://doi.org/10.1016/j.cortex.2009.08.007>
- Friedmann, N., & Gvion, A. (2003). Sentence comprehension and working memory limitation in aphasia: A dissociation between semantic-syntactic and phonological reactivation. *Brain and Language*, 86(1), 23–39. [https://doi.org/https://doi.org/10.1016/S0093-934X\(02\)00530-8](https://doi.org/https://doi.org/10.1016/S0093-934X(02)00530-8)
- Furumoto, H. (2006). Pure misallocation of “0” in number transcoding: a new symptom of right cerebral dysfunction. *Brain and Cognition*, 60(2), 128–138. <https://doi.org/10.1016/j.bandc.2005.10.002>
- García-Orza, J., & Perea, M. (2011). Position coding in two-digit arabic numbers. *Experimental Psychology*, 58(2), 85–91. <https://doi.org/10.1027/1618-3169/a000071>
- Göbel, S. M. (2020). *Number writing as a longitudinal predictor of arithmetic development and early fraction understanding*. Presented at the 3rd conference of the Mathematical Cognition and Learning Society.
- 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. <https://doi.org/10.1016/j.jecp.2013.10.001>
- González, E. G., & Kolars, P. A. (1982). Mental manipulation of arithmetic symbols. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 8(4), 308–319. <https://doi.org/10.1037/0278-7393.8.4.308>
- Granà, A., Lochy, A., Girelli, L., Seron, X., & Semenza, C. (2003). Transcoding zeros within complex numerals. *Neuropsychologia*, 41(12), 1611–1618. [https://doi.org/10.1016/S0028-3932\(03\)00109-X](https://doi.org/10.1016/S0028-3932(03)00109-X)
- Gvion, A., & Friedmann, N. (2012). Phonological short-term memory in conduction aphasia. *Aphasiology*, 26(3–4), 579–614. <https://doi.org/10.1080/02687038.2011.643759>
- Hanks, T. D., Mazurek, M. E., Kiani, R., Hopp, E., & Shadlen, M. N. (2011). Elapsed decision time affects the weighting of prior probability in a perceptual decision task. *Journal of Neuroscience*, 31(17), 6339–6352. <https://doi.org/10.1523/JNEUROSCI.5613-10.2011>
- Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science*, 298(5598), 1569–1579. <https://doi.org/10.1126/science.298.5598.1569>
- Hayek, M., Karni, A., & Eviatar, Z. (2020). Transcoding number words by bilingual speakers of Arabic: writing multi-digit numbers in a units-decades inverting language. *Writing Systems Research*, 1–15. <https://doi.org/10.1080/17586801.2020.1787298>
- Hung, Y. H., Pallier, C., Dehaene, S., Lin, Y.-C., Chang, A., Tzeng, O. J.-L., & Wu, D. H. (2015). Neural correlates of merging number words. *NeuroImage*, 122, 33–43. <https://doi.org/https://doi.org/10.1016/j.neuroimage.2015.07.045>
- Kallai, A. Y., & Tzelgov, J. (2012). The place-value of a digit in multi-digit numbers is processed automatically. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38(5), 1221–1233. <https://doi.org/10.1037/a0027635>
- Kording, K. P., & Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning. *Nature*, 427(6971), 244–247. <https://doi.org/10.1038/nature02169>

- Lochy, A., Domahs, F., Barthä, L., & Delazer, M. (2004). Specific order impairment in arabic number writing: A case-study. *Cognitive Neuropsychology*, 21(5), 555–575. <https://doi.org/10.1080/02643290342000618>
- Lochy, A., Pillon, A., Zesiger, P., & Seron, X. (2002). Verbal structure of numerals and digits handwriting: New evidence from kinematics. *The Quarterly Journal of Experimental Psychology Section A*, 55(1), 263–288. <https://doi.org/10.1080/02724980143000271>
- Mahowald, K., James, A., Futrell, R., & Gibson, E. (2016). A meta-analysis of syntactic priming in language production. *Journal of Memory and Language*, 91, 5–27. <https://doi.org/10.1016/j.jml.2016.03.009>
- Malone, S. A., Pritchard, V. E., & Hulme, C. (2021). Separable effects of the approximate number system, symbolic number knowledge, and number ordering ability on early arithmetic development. *Journal of Experimental Child Psychology*, 208, 105120. <https://doi.org/https://doi.org/10.1016/j.jecp.2021.105120>
- Marangolo, P., Nasti, M., & Zorzi, M. (2004). Selective impairment for reading numbers and number words: A single case study. *Neuropsychologia*, 42(8), 997–1006. <https://doi.org/10.1016/j.neuropsychologia.2004.01.004>
- McCloskey, M. (1992). Cognitive mechanisms in numerical processing: Evidence from acquired dyscalculia. *Cognition*, 44(1–2), 107–157. [https://doi.org/10.1016/0010-0277\(92\)90052-J](https://doi.org/10.1016/0010-0277(92)90052-J)
- McCloskey, M., Caramazza, A., & Basili, A. (1985). Cognitive mechanisms in number processing and calculation: Evidence from dyscalculia. *Brain and Cognition*, 4(2), 171–196. [https://doi.org/10.1016/0278-2626\(85\)90069-7](https://doi.org/10.1016/0278-2626(85)90069-7)
- McCloskey, M., Sokol, S. M., & Goodman, R. A. (1986). Cognitive processes in verbal-number production: Inferences from the performance of brain-damaged subjects. *Journal of Experimental Psychology: General*, 115(4), 307–330. <https://doi.org/10.1037/0096-3445.115.4.307>
- Mehler, J., & Carey, P. (1967). Role of surface and base structure in the perception of sentences. *Journal of Verbal Learning and Verbal Behavior*, 6(3), 335–338. [https://doi.org/10.1016/S0022-5371\(67\)80122-1](https://doi.org/10.1016/S0022-5371(67)80122-1)
- Miura, I. T., & Okamoto, Y. (1989). Comparisons of U.S. and Japanese first graders' cognitive representation of number and understanding of place value. *Journal of Educational Psychology*, 81(1), 109–114. <https://doi.org/10.1037/0022-0663.81.1.109>
- Moeller, K., Pixner, S., Kaufmann, L., & Nuerk, H. C. (2009). Children's early mental number line: Logarithmic or decomposed linear? *Journal of Experimental Child Psychology*, 103(4), 503–515. <https://doi.org/10.1016/j.jecp.2009.02.006>
- 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(5), 1837–1851. <https://doi.org/https://doi.org/10.1016/j.ridd.2011.03.012>
- 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(3), 707–727. <https://doi.org/10.1016/j.jecp.2013.07.008>
- Noël, M. P., & Seron, X. (1993). Arabic number reading deficit: A single case study or when 236 is read (2306) and judged superior to 1258. *Cognitive Neuropsychology*, 10(4), 317–339. <https://doi.org/10.1080/02643299308253467>
- Noël, M. P., & Seron, X. (1997). On the existence of intermediate representations in numerical processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23(3), 697–720. <https://doi.org/10.1037/0278-7393.23.3.697>
- Nuerk, H. C., & Willmes, K. (2005). On the magnitude representations of two-digit numbers. *Psychology Science*, 47(1), 52–72.
- Perry, M. (2019). *Simanneal: Python module for simulated annealing optimization*. <http://github.com/perrygeo/simanneal>
- Pickering, M. J., & Ferreira, V. S. (2008). Structural priming: A critical review. *Psychological Bulletin*, 134(3), 427–459. <https://doi.org/10.1037/0033-2909.134.3.427>
- 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(6), 2683–2689. <https://doi.org/10.1016/j.ridd.2011.06.004>

- Pourquié, M., & Nespoulous, J.-L. (2018). On linguistic properties of verbal number systems: A cross-linguistic study of number transcoding errors observed in a Basque–French bilingual patient with aphasia. *Lingua*, 203, 27–35. <https://doi.org/10.1016/j.lingua.2017.10.002>
- Power, R., & Dal Martello, M. F. (1990). The dictation of Italian numerals. *Language and Cognitive Processes*, 5(3), 237–254. <https://doi.org/10.1080/01690969008402106>
- Power, R., & Dal Martello, M. F. (1997). From 834 to Eighty Thirty Four: The reading of Arabic numerals by seven-year-old children. *Mathematical Cognition*, 3(1), 63–85. <https://doi.org/10.1080/135467997387489>
- Proios, H., Weniger, D., & Willmes, K. (2002). Number representation deficit: a bilingual case of failure to access written verbal numeral representations. *Neuropsychologia*, 40(13), 2341–2349. [https://doi.org/https://doi.org/10.1016/S0028-3932\(02\)00085-4](https://doi.org/https://doi.org/10.1016/S0028-3932(02)00085-4)
- Qasim-Masarwa, H., Marsel-Levi, M., & Dotan, D. (2020). *How do we write numbers: left-to-right digit, or first-to-last word?* http://mathinklab.org/posters/QasimLeviDotan2020_ReversedNumberWritingInArabic.pdf
- R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <http://www.r-project.org>
- Seron, X., & Fayol, M. (1994). Number transcoding in children: A functional analysis. *British Journal of Developmental Psychology*, 12(3), 281–300. <https://doi.org/10.1111/j.2044-835X.1994.tb00635.x>
- Summerfield, C., & de Lange, F. P. (2014). Expectation in perceptual decision making: Neural and computational mechanisms. *Nature Reviews Neuroscience*, 15(11), 745–756. <http://dx.doi.org/10.1038/nrn3838>
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157), 1124–1131. <https://doi.org/10.1126/science.185.4157.1124>
- Xenidou-Dervou, I., Molenaar, D., Ansari, D., van der Schoot, M., & van Lieshout, E. C. D. M. (2017). Nonsymbolic and symbolic magnitude comparison skills as longitudinal predictors of mathematical achievement. *Learning and Instruction*, 50, 1–13. <https://doi.org/https://doi.org/10.1016/j.learninstruc.2016.11.001>
- Yuan, L., Prather, R. W., Mix, K. S., & Smith, L. B. (2019). Preschoolers and multi-digit numbers: A path to mathematics through the symbols themselves. *Cognition*, 189, 89–104. <https://doi.org/https://doi.org/10.1016/j.cognition.2019.03.013>
- 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(1), 60–77. <https://doi.org/10.1016/j.jecp.2008.04.003>