How does the syntax of counting affect learnability? Evidence from artificial language learning.

Sebastian Holt, David Barner
Department of Psychology, UC San Diego
{sholt, dbarner}@ucsd.edu

Abstract

Generative syntax and a consistently ordered count routine are both understood to have central roles in learning a number system. However, there has been little experimental exploration of how the diversity of each of these features alters the inductive landscape of number learning, as most empirical work has been constrained to correlational studies. We present a causal manipulation both of syntactic structures and counting procedures, using an artificial language paradigm. Our findings suggest that (1) learners have a greater facility with conjunctive over multiplicative rules of composition, (2) counting procedures help learners to recall words independently of syntax, and (3) predictable syntax helps learners to use numerical concepts, independently of - and possibly despite - counting routines.

Keywords: number; artificial languages; counting; syntax

Introduction

Humans are able to use concepts of large, precise magnitude, such as *2023*, using culturally transmitted verbal number systems. However, these systems show a remarkable diversity in both (1) the syntactic structures that they exhibit (e.g., their *base systems*) and (2) the ways that they are used (e.g., their *counting procedures*), presenting learners with very different challenges as they acquire different number systems. For example, in a base-10 system like Mandarin, learners have to memorize a base set of ten lexical items (e.g., 1-10), which are then composed using syntactic rules to represent larger numbers (e.g., 91). Most learners of base-10 number systems, like Mandarin or English, also rehearse these number words in a *counting procedure*, which involves repeating numbers in an ordered sequence, such as "1, 2, 3…", and which is usually memorized by children long before they understand the meaning of any number words (Fuson, 1988).

Number systems feature a range of different base systems from base-2 to base-20 (Hammarström, 2010; Comrie, 2011), requiring learners to memorize a smaller or larger number of lexical primitives and rules of composition in order to represent the same set of cardinalities. Furthermore, people do not always engage in counting to learn them. For example, the Western Tribe of Torres Straits had a number system featuring the words *urapun* (1), *okosa* (2), *okosa urapun* (3), *okosa okosa* (4), *okosa okosa urapun* (5), and *okosa okosa okosa* (6) (Haddon, 1890). This system featured a clear base-2 with one additive rule, and in this sense was a counting system, but was not learned via a counting procedure. Instead, these words were used to refer to quantities when they needed to be communicated, but they were not enumerated in a structured procedure; much like the non-exact English quantifiers 'a couple' or 'a few'.

How might such diversity in counting systems impact the acquisition of number words and their meanings? Accounts of how children acquire number vary with respect to the role they posit for syntactic structures and counting procedures. For example, according to Carey (2004, 2009), children learn the meanings of large number words based on an analogy to the linear order of the count list. By this account, children begin the learning process by associating the small number words "one", "two", and "three" with representations of small sets, but learn larger number word meanings by observing that for every item added to a set, it can be described by counting up one word in the count list (see also Gentner, 2010; Sarnecka & Carey, 2008; Carey & Barner, 2019). Completing this analogy between the ordered set of cardinalities and the sequence of the count list then supports a broader inductive inference, that every number word, N, has a successor whose value is $N+1$, such that there exists an infinite number of numbers. However, on this account, the syntax of counting itself plays no significant role in how or when this inductive inference is made.

Other accounts also hypothesize that the linear structure of the count list plays a role in number word learning, but argue that the inductive inference posited by Carey does not occur until much later, and may depend on additional, more subtle, aspects of how counting is structured. For example, children learning Cantonese, which uses highly predictable base-10 rules to generate numbers (e.g., 22 translates as *2-10-2*) are faster to learn their verbal count list than children learning a very irregular counting system like Hindi, which features many irregular forms that are difficult to predict from rules (Ho & Fuson, 1998; Miller & Stigler, 1987; Miller et al., 1995; Schneider et al., 2020). Critically, children learning predictable counting systems are quicker to learn that every number has a successor, and their knowledge of counting

In M. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the 45th Annual Conference of the Cognitive Science Society*.

rules is related to the belief that numbers are infinite (Cheung et al., 2017; Chu et al., 2020; Guerrero et al., 2020; Miller et al., 1995; Schneider et al., 2020; Miura et al., 1993; Okamoto et al., 1993; Miura & Okamoto, 2013; Song & Ginsburg, 1987). Together, these studies suggest that children's ability to extract the syntactic rules that govern counting may play a role in their discovery of how counting represents number.

Although past studies suggest that cross-cultural differences in counting practices may impact learning, these reports are almost exclusively correlational in nature, leaving open a range of alternative explanations for reported effects. Further, previous studies have been limited to examining variability across base-10 systems, which differ mainly in the relative regularity of their rule systems. Although some studies have probed children's ability to notice and exploit the morphosyntactic composition of number words using experimental methods (Barrouillet et al., 2010; Cheung et al., 2016), they did not contrast different counting systems to assess the role of variability, including the size of the memorized base, types of compositional rules (e.g., additive vs. multiplicative), and ways in which learners encounter number words (e.g., everyday language or counting). Further, past studies have not investigated whether learners are faster to acquire a counting system if it is presented to them as a rote, ordered, counting sequence, or if they can learn a system as easily if number words are presented haphazardly in speech, without a fixed routine or particular order.

In the present study, we took an experimental approach to test the role of syntactic structures and procedures in the acquisition of counting systems. To do so, we trained adult learners in two experiments on a range of different counting systems in an artificial language learning paradigm. In Experiment 1, participants were trained to count in a number system exhibiting one of six different numerical bases. We found that while base systems differed little, learners have a greater facility with conjunctive over multiplicative rules of composition. In Experiment 2, participants learned number systems that either were rehearsed in a consistently ordered count routine, exhibited a compositional semantics, or both. We found that rehearsed order and compositional rules facilitated word-recall and reasoning about number concepts, respectively.

Experiment 1: Base Systems

In Experiment 1, participants learned six artificial number systems that differed only in the size of their numerical base, requiring them to memorize a different number of lexical primitives and compositional rules (e.g., counting to 4 in a base-3 system requires three lexical primitives and one additive rule, whereas counting to 4 in a base-4 system requires memorizing four lexical primitives but no rules).

With this task, we asked (1) whether some numerical bases were more or less learnable than others due to this tradeoff between words and rules, and (2) how participants reason about novel numbers they had not been trained on. In Experiment 1a we tested bases 2, 3, 4, and 5, and in Experiment 1b we added the bases 8 and 10.

Methods

Participants Based on a pre-registered power analysis, 235 adult participants were recruited online via Prolific. 53 were excluded due to our pre-registered criteria, e.g., low accuracy. In Experiment 1a, *n*=31, 30, 32, & 30 participants were recruited for base-2, 3, 4, & 5, respectively. In Experiment 1b, 30 were recruited for base-8, and 29 for base-10. Compensation was targeted at $14/hour, though usually exceeded this with performance bonus (*m* = $1.90).

Stimuli Artificial number systems for each game combined lexical atoms via a compositional syntax. Atoms were randomly generated from consonant-vowel bigrams (e.g., *ka*) such that no consonant or vowel was used in more than one atom per number system. Apart from the size of the base, all systems in Experiments 1a and 1b had identical syntactic

A

Lexical Primitives

1 = *no*; 2 = *ka*; 3 = *hi*; 9 = *se*

1-10: *no, ka, hi, hino, hika, kahi, kahino, kahika, se, seno*

Rule	Artificial Example	English Example
Add. +	4 is *hi-no* (3+1)	*Fifteen* = 5 + 10
Mult. ×	6 is *ka-hi* (2*3)	*Twenty* = 2 × 10
Exp. ^	9 (3^2) is a primitive	*Thousand* = 10^3

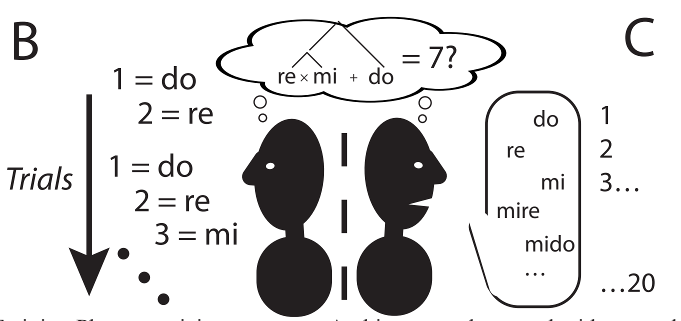

Figure 1. A. An example artificial base-3 number system. **B.** In the Training Phase, participants saw an Arabic numeral equated with a novel number word, and typed it. Trials appeared in numerical order, and the sequence repeated with every novel word. **C.** In the Generalization Phase, participants had to recall the count list they had trained with, and infer the syntax to label the numbers 11-15 (Experiment 1a) or 11-20 (1b).

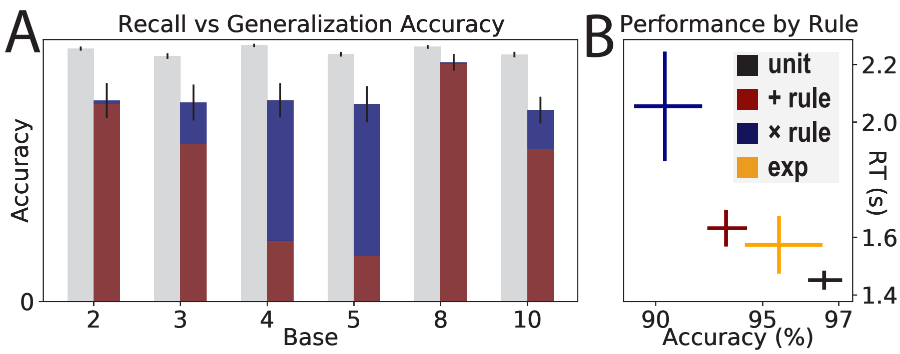

Figure 2. A. Recall (grey, left bars) and generalization (red, blue, right bars) accuracy for each base size of Experiment 1. Red & blue signify additive & multiplicative strategies used by participants. All errorbars are 95-CIs. **B.** Reaction time and accuracy, during recall, for target numbers featuring each rule type: atoms (black), additive (red), multiplicative (blue), exponents (gold). Errors are 95-CIs.

rules (**Figure 1A**): For each base, b, (1) atomic unit words denoted quantities up to b, (2) quantities larger than b but indivisible by it used addition, where unit atoms were added to the end of the word to express the remainder, (3) quantities of $2 \times b$ or greater used multiplication, where the largest compatible unit was placed before the b (like digit-multiplier constructions, such as *six-dozen*, 6*12, in English), and (4) quantities of $b \times b$ or greater used an exponential atom (like *hundred*, 10^2), composed like the unit terms. There were six bases: 2, 3, 4, 5, 8, and a base-10 (really, *no base*) system with 10 atoms and no rules.

Training Phase After reading an instructions page and passing a comprehension quiz, participants entered the **training phase**, during which they rehearsed a count list, in ascending order, of 10 artificial number words (**Figure 1B**). Each participant learned number words belonging to one of the six base systems, depending on condition. The first time they encountered a number word, they were explicitly taught its meaning (e.g., the screen read "1 = *ka*"), and they were asked to type the number word in an input field below. Subsequently, every time they were cued with that cardinality, they had to recall the number word from memory (e.g., the screen read "1 = ?"). Before each new numeral was introduced, the participant was tested on all previously trained numbers: First, they were trained and tested on the words for 1 and 2, then they were trained on 3 and tested on 1, 2, and 3, etc. until they reached 10. In this way, exposure to each number corresponded roughly to their frequency in natural language (e.g., BNC Consortium, 2007). At the end of this process, the participant rehearsed the count list one last last time, from 1-10. Responses for these 64 **recall trials** (all responses in the training phase except the first exposure) were used below to assess the relative learnability of the different base systems, in addition to data from the Generalization Phase. To motivate accurate responses, participants received a bonus for accuracy: \$0.03 per correct guess on the first try, \$0.02 if correct on the second try, and \$0.01 correct on the third try. On submitting a correct or third try, they were advanced to the next word in the sequence. Participants were instructed to complete trials "as quickly and accurately as possible", but were not timed.

Generalization Phase After completing the Training Phase of the experiment, participants completed the Generalization Phase, wherein they were shown a chart with Arabic numerals from 1-15 (or 1-20, in Experiment 1b) in the left-hand column, and empty text-entry fields on the right. Participants entered number words from the artificial language in the text entry fields, one at a time. Answers for each number could not be edited after being entered. The last five or ten numbers, from 10-15 or 20, were an opportunity for participants to generalize the number system to novel cardinalities that they had not yet encountered.

Results

Rule types may matter more than base size. To test whether base size impacted the learnability of a counting system, we created a measure of response time (consisting of reaction time until the first keystroke, plus the time interval during which participants were typing divided by the length of the text they entered) and also measured accuracy on quiz trials. Response time analyses were repeated on reaction time, to similar qualitative results. We constructed linear mixed effects regressions predicting each of these variables based on base-size, with random intercepts for each participant. We found no effect of base size on our normed response time measure in 1a or 1b (small bases: $b = 22.29$ ms, $t = .307$, $p = .759$; large bases: $b = -26.61$, $t = -.741$, $p = .46$). While we found a negative effect of base size on accuracy in Experiment 1b (large bases: $b = -.223$, $z = -3.103$, $p = .002$), this did not hold for Experiment 1a (small bases; $b = -.005$, $z = -.046$, $p = .963$), or Experiment 1 as a whole ($b = -.012$, $z = -.315$, $p = .753$; **Figure 2A**). These results are hard to interpet, however, as ceiling effects from overtraining may explain the lack of variability.

While our analysis of base size *per se* was inconclusive, the different types of rules that learners of each base were exposed to may have had subtle and conflicting effects that did not correlate directly with base-size (e.g., multiplicative rules were common in base-4, but not base-2 or base-8). In an expanded regression, we coded each number word that participants had to learn as the composition of several distinct rules (**Figure 1A**), in addition to its membership in a base system and word-length, measured in syllables. These rules were (1) additive (2) multiplicative rules of composition, as well as (3) the presence of a lexical atom with exponential meaning (i.e., 4 in the base-2 condition and 9 in base-3). Any one cardinality might be composed from different rules, depending on the base: e.g., 3 in base-3 is composed only of a lexical primitive, whereas in base-2 it features an additive rule composing 2+1. We tested this model on the entire dataset of Experiment 1, and found that only multiplicative constructions were significantly related to both increased response time (b = 580.21, t = 3.89, $p < .001$), and lower accuracy (b = -.929, t = -1.966, p = 0.049; **Figure 2B**), suggesting that multiplicative compositions may be harder to learn than additive rules or than memorizing syllables.

10-item count lists with different bases imply different compositional rules. While trials in the Training Phase tested participants' ability to recall words they had just learned, they did not address how participants mentally represented the rules of each system. For example, although rules were available to be learned in most systems, learners could nevertheless ignore these rules and learn the number words by rote, or create an alternative set of rules to the ones we had used to generate the artificial languages. Indeed, we found that participants rarely learned the full set of additive, multiplicative, and exponential rules that had produced the counting system they learned. Their labels were identical to our predicted generalizations 38% of the time. For each base, these figures were 36% (base-*2*), 24% (*3*), 33% (*4*), 49% (*5*), 42% (*8*), and 38% (*10*). We therefore also measured accuracy using more permissive additive or multiplicative parses. Additive parses summed the meanings of all syllables in a word, regardless of order. Multiplicative parses had this, and one extra grammatical rule: whenever a small number preceded a larger number, it was interpreted as a multiplicative digit-multiplier phrase, analogous to natural languages (e.g., English *two-hundred*, 2×100). In addition to these directly compositional strategies, either strategy might be employed to recursively generate each number based on its predecessor in the count list. For example, if a participant skipped one number in the count list but continued with labels that increased by +1 via additive or multiplicative grammars, then subsequent labels would be correct. Thus parsed, participants' accuracy was higher, though still below the training phase (.793, 95%-*CI*: [.77, .815]; **Figure 2A**). Also, while multiplicative rules were used in base-4 and 5 conditions to label the target number, additive rules were more common in others, accounting for 75% of correct labels. Participants used additive rules more frequently than our generative rule system predicted, ($\chi^2(3, N = 2410 \text{ trials}) = 143.625$, $p < 0.001$), suggesting again that multiplicative rules may be less intuitive, despite being observed and used by learners.

Experiment 2: Count Procedures

In Experiment 1, we investigated the learnability of different base systems, all of which were trained as part of a list, ordered in terms of magnitude. On some accounts, the presence of this ordering is important to learning the meanings of number words, since their meanings are defined by their ordinal relations to one another (Carey, 2004). On other accounts, exposure to an ordered sequence may be important because it provides exposure to the recursive syntax of number words, making the presence of compositional rules more obvious. Finally, it is possible that although children in the west are typically taught number words as part of an ordered procedure, this ordering is not actually essential to learning their meanings, or extracting rules for creating new numbers. Instead, learners may be able to learn these rules just as they acquire many other grammatical rules (e.g., via cross-situational learning that includes many distinct tokens of words across different utterances in their analysis) (Gleitman 1990; Pinker 1994).

In Experiment 2, we asked whether three groups of participants would learn a number system as effectively when presented with number systems featuring (1) compositional syntax and a consistently ordered count list, (2) compositional syntax but no consistent order of exposure, and (3) *no* compositional syntax, but a consistently ordered count list. Also, we included a magnitude comparison task to assess whether participants had mapped number words to conceptual content.

Methods

Participants 119 adult participants were recruited online via Prolific, of whom 28 were excluded due to our pre-registered criteria, e.g., low accuracy. Samples were N=30, 31, and 30 in the **rules+order**, **order-only**, and **rules-only** conditions, respectively. Compensation was targeted at \$14/hour + performance bonus (m = \$2.36).

Stimuli Artificial number words were generated as in the base-5 condition of Experiment 1. In the **order-only** condition, the mapping between number words and meanings was randomly shuffled.

Training Phase. Each participant was taught an artificial number system in a similar manner to Experiment 1. Two key features distinguish this experiment from the previous. First, all groups learned a base-5 system, selected because it exhibits combinatorial structure over half the count list, and requires only additive and multiplicative rules (no exponents) to label any number up to 20 (the highest cardinality appearing in the experiment). Second, in order to independently manipulate the presence of combinatorial rules and an ordered count procedure, we divided participants into three conditions. In Condition 1, (the "**rules+order**" condition) number words both featured compositional rules, and were trained as a consistent, ordered list, as in Experiment 1, which was rehearsed in a count list that grew by +1 with every iteration. In Condition 2 ("**order-only**"), words were trained in an ordered list, but the words generated via compositional rules were remapped to an arbitrary, different, cardinality within the count list. Thus, if the lexicon learned in the **rules+order** condition was {1 = *do*, 2 = *re*, 3 = *mi*, 4 = *mido*, 5 = *mire*, etc.}, the lexicon learned in the **order-only** condition might be {1 = *mire*, 2 = *mi*, 3 = *do*, 4 = *re*, 5 = *mido*, etc.}. In Condition 3 ("**rules-only**"), the lexicon was identical to the **rules+order** condition, but the words were trained in a random order (i.e., not in order of magnitude). Critically, the frequency of each number word was matched with the synonymous word in the control condition, such that the word for 1 was most frequent, followed by the word meaning 2, then 3, etc.

Generalization Phase. The generalization phase was identical to the one used in Experiment 1b. Novel number values to be filled in by the participant ranged from 11-20.

Magnitude Comparison. In each trial, participants were presented with a pair of unequal number words from the count list they had just learned, and asked to click on the one that represented the 'bigger' value. Each trial had a time limit of 4 seconds, during which a progress bar at the top of the screen shrank to indicate remaining time. Over the task, all 45 unequal combinations of number words were queried.

Results

Counting helps number word recall. To test the relative contributions of rules and order on *recalling* learned words, we constructed mixed effects linear regressions predicting

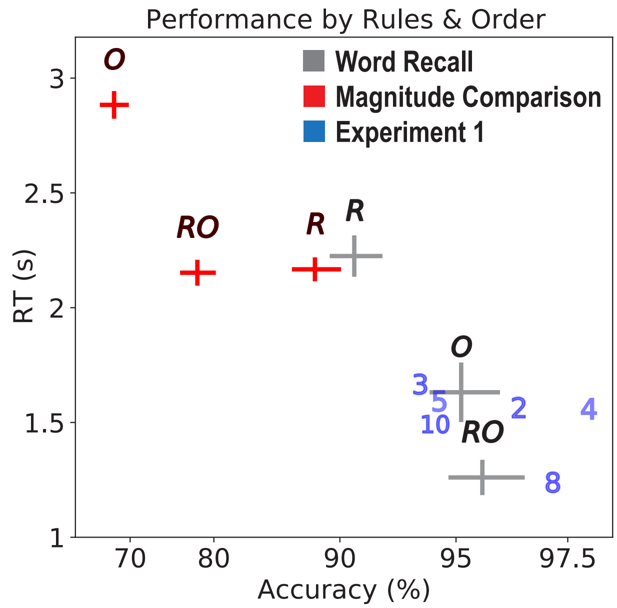

Figure 3. Reaction time and accuracy, during recall (grey) and magnitude comparison (red) trials of Experiment 2, for each condition (***R*** = rules-only; ***O*** = order-only; ***RO*** = rules+order). Means for each base in Experiment 1 for comparison (blue). Errors are 95-CIs.

accuracy and response time from task condition, target number, cumulative frequency of that number, an interaction between condition and frequency, and random intercepts per participant. Participants in the **rules-only** condition were significantly slower ($b = 1404.78$, $t = 5.493$, $p < .001$) and less accurate ($b = -1.57$, $z = -3.238$, $p = .001$) in recalling artificial number words than in the **rules+order** condition. Participants in the **order-only** condition were also slightly slower than **rules+order** ($b = 688.78$, $t = 2.716$, $p = .008$), but did not differ in accuracy ($b = -.469$, $z = -.945$, $p = .345$), (**Figure 3**, grey). The similarity between the **rules+order** and **order-only** conditions suggests that when learners benefit from learning number words in an ordered sequence, transparent rules within the system may not greatly facilitate word recall - and are much less likely to be sufficient by themselves.

Compositional rules help learn number word meanings. To test the contributions of rules and order on learning the *meanings* of words, we constructed mixed effects regressions to predict accuracy and response time on the magnitude comparison task. These included terms for task condition, the sum of the compared numbers (a+b), the numerical distance between them (a-b), the difference in the length of the words (number of syllables), interactions between word-length & distance, word-length & condition, and as intercepts for each participant. Participants in the **order-only** condition were significantly slower at selecting the greater of two number words ($b = 392.01$, $t = 2.338$, $p = .021$) and less accurate ($b = -.884$, $z = -2.131$, $p = .033$) than participants in other conditions. In contrast, participants in the **rules-only** condition showed comparable response latencies to the **rules+order** ($b = -47.28$, $t = -.28$, $p = 0.78$), but were significantly more accurate in their responses ($b = 0.882$, $z = 2.036$, $p = .042$; **Figure 3**, red), suggesting that compositional rules are uniquely helpful for learning the meanings of number words. To test whether this advantage was due to a low-level word-length heuristic or specifically to numerical distance, we compared the full model to one with word-length removed. We found that while distance was no longer significantly predictive of reaction time when word-length was added to the model ($b = -3.15$, $t = -.233$, $p = .816$), it remained significantly predictive of accuracy ($b = 0.455$, $z = 11.459$, $p < .001$)[1]. These analyses found that exposure to compositional syntax facilitates learning the meaning of number words independently of rehearsing those number words in an ordered sequence, suggesting that the same mechanisms that allow humans to acquire non-numerical grammatical constructions, such as the past-tense, may be sufficient for learning a generative number system. Further, as the participants who learned a syntax with no consistent order (**rules-only**) outperformed those who learned a syntax in a consistent order (**rules+order**), learners may even be better motivated to detect and remember syntactic rules in the absence of a count routine. This stands in marked contrast to the Training Phase of the experiment, in which rehearsal of an ordered count list was important for success, while exposure to compositional syntax was not.

Discussion

We manipulated the syntactic structures and counting procedures available to learners of artificial number systems, and found several main results. First, in Experiment 1 we found that the size of a numerical base did not impact how easily different number systems were learned. Second, we found that although base size itself did not have an impact on learning, the nature of the rules that were implicated by different bases did. In particular, while additive rules of composition (e.g., 3+1 to represent 4) were relatively easy for participants to learn and generalize to untrained numbers, multiplicative rules posed a greater challenge. Third, in Experiment 2 we found that learning number words in a consistently ordered count list helped participants to recall those words, but not to generalize the number system they had learned. Finally, learning a number system with a transparent, regular syntax helped participants learn the meanings of number words, even though it did not facilitate recall of words.

In this study, we trained numerate adults to probe the impacts of counting structures and procedures on learning. While there is a long tradition of training adult learners on artificial grammars to assess learnability, future studies should consider whether naive learners, who have not previously acquired a number system, might show different learning strategies. Studies now in progress are exploring this question, using methods adapted from the experiments presented here to study young semi-numerate children who are just learning to use English number words. In doing so, this work will build on recent studies investigating children's understanding of numerical composition (Cheung et al., 2016; Park et al., 2022), to explore questions of cognitive development that are difficult to assess using observation of naturalistic language. development

Also of interest is the role that communicative processes play in the transmission and evolution of counting systems. In our studies, learning occurred via a strict pedagogical process in which there was one teacher (the software) and one learner (the participant). However, in the wild counting systems have arisen via a multigenerational transmission process, in which each generation learns and transmits their practices resulting in irregularities, systems with mixed bases, and partial generalizations. Future studies should examine how multigeneration communicative transmission chains impact the form that counting systems take, and whether diachronic processes of change favor particular bases, rules, or procedures over others. Such work offers the possibility of uniting work on the learnability of counting systems with existing theories of efficient communication (e.g., Gibson et al., 2019; Xu, Liu, & Regier, 2020) while using methods previously deployed for studying the emergence and change of symbolic systems (Kirby et al., 2014; Hawkins et al., 2021).

OSF pre-registration:
https://osf.io/rwqk7/

All code and materials available at:
https://github.com/SebastianHolt/number_syntax

[1] This finding may be corroborated by a similar pattern in the Generalization phase, analyses of which are forthcoming in the Github repository.

Acknowledgements

Many thanks to Michael Ho, Elisabeth Marchand, and Charis Holland for their hard work, and members of the Language Development Lab, and Cognitive Tools Lab at UC San Diego for insightful feedback throughout.

References

Barrouillet, P., Thevenot, C., & Fayol, M. (2010). Evidence for knowledge of the syntax of large numbers in preschoolers. *Journal of Experimental Child Psychology, 105*(3), 264-271.

BNC Consortium. (2007). British national corpus. *Oxford Text Archive Core Collection*.

Carey, S. (2004). Bootstrapping & the origin of concepts. *Daedalus, 133*(1), 59-68.

Carey, S. (2009). Where our number concepts come from. *The Journal of philosophy, 106*(4), 220-254.

Carey, S., & Barner, D. (2019). Ontogenetic origins of human integer representations. *Trends in cognitive sciences, 23*(10), 823-835.

Cheung, P., Dale, M., & Le Corre, M. (2016). A cross-linguistic investigation on the acquisition of complex numerals. In *CogSci* (pp. 2711-2716).

Cheung, P., Rubenson, M., & Barner, D. (2017). To infinity and beyond: Children generalize the successor function to all possible numbers years after learning to count. *Cognitive psychology, 92*, 22-36.

Chu, J., Cheung, P., Schneider, R. M., Sullivan, J., & Barner, D. (2020). Counting to Infinity: Does learning the syntax of the count list predict knowledge that numbers are infinite?. *Cognitive Science, 44*(8), e12875.

Comrie, B. (2011). Typology of numeral systems. Numeral types and changes worldwide. *Trends in Linguistics. Studies and monographs, 118*.

Fuson, K. C. (1988). Children's Counting and Concepts of Number Springer-Verlag: New York. *FusonChildren's counting and concepts of number1988*.

Gentner, D. (2010). Bootstrapping the mind: Analogical processes and symbol systems. *Cognitive Science, 34*(5), 752-775.

Gibson, E., Futrell, R., Jara-Ettinger, J., Mahowald, K., Bergen, L., Ratnasingam, S., ... & Conway, B. R. (2017). Color naming across languages reflects color use. *Proceedings of the National Academy of Sciences, 114*(40), 10785-10790.

Gibson, E., Futrell, R., Piantadosi, S. P., Dautriche, I., Mahowald, K., Bergen, L., & Levy, R. (2019). How efficiency shapes human language. *Trends in cognitive sciences, 23*(5), 389-407.

Gleitman, L. (1990). The structural sources of verb meanings. *Language acquisition, 1*(1), 3-55.

Guerrero, D., Hwang, J., Boutin, B., Roeper, T., & Park, J. (2020). Is thirty-two three tens and two ones? The embedded structure of cardinal numbers. *Cognition, 203*, 104331.

Haddon, A. C. (1890). The ethnography of the western tribe of Torres Straits. *The Journal of the Anthropological Institute of Great Britain and Ireland, 1*9, 297-440.

Hammarström, H. (2010). Rarities in numeral systems. Rethinking universals: *How rarities affect linguistic theory, 45*, 11-53.

Hawkins, R. D., Franke, M., Frank, M. C., Goldberg, A. E., Smith, K., Griffiths, T. L., & Goodman, N. D. (2022). From partners to populations: A hierarchical Bayesian account of coordination and convention. *Psychological Review*.

Ho, C. S. H., & Fuson, K. C. (1998). Children's knowledge of teen quantities as tens and ones: Comparisons of Chinese, British, and American kindergartners. *Journal of Educational Psychology, 90*(3), 536.

Kirby, S., Griffiths, T., & Smith, K. (2014). Iterated learning and the evolution of language. *Current opinion in neurobiology, 28*, 108-114.

Miller, K. F., & Stigler, J. W. (1987). Counting in Chinese: Cultural variation in a basic cognitive skill. *Cognitive development, 2*(3), 279-305.

Miller, K. F., Smith, C. M., Zhu, J., & Zhang, H. (1995). Preschool origins of cross-national differences in mathematical competence: The role of number-naming systems. *Psychological Science, 6*(1), 56-60.

Miura, I. T., Okamoto, Y., Kim, C. C., Steere, M., & Fayol, M. (1993). First graders' cognitive representation of number and understanding of place value: Cross-national comparisons: France, Japan, Korea, Sweden, and the United States. *Journal of educational psychology, 85*(1), 24.

Miura, I. T., & Okamoto, Y. (2013). Language supports for mathematics understanding and performance. In *The development of arithmetic concepts and skills* (pp. 251-264). Routledge.

Pinker, S. (1994). How could a child use verb syntax to learn verb semantics?. *Lingua, 92*, 377-410.

Sarnecka, B. W., & Carey, S. (2008). How counting represents number: What children must learn and when they learn it. *Cognition, 108*(3), 662-674.

Schneider, R. M., Sullivan, J., Marušič, F., Biswas, P., Mišmaš, P., Plesničar, V., & Barner, D. (2020). Do children use language structure to discover the recursive rules of counting?. *Cognitive psychology, 117*, 101263.

Song, M. J., & Ginsburg, H. P. (1987). The development of informal and formal mathematical thinking in Korean and US children. *Child development*, 1286-1296.

Xu, Y., Liu, E., & Regier, T. (2020). Numeral systems across languages support efficient communication: From approximate numerosity to recursion. *Open Mind, 4,* 57-70.