UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

Artificial Languages and Cognitive Constraints on Symbolic Number

Permalink

https://escholarship.org/uc/item/3xf465n5

ISBN

9798293839292

Author

Holt, Sebastian Martin Newman

Publication Date

2025-09-16

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA SAN DIEGO

Artificial Languages and Cognitive Constraints on Symbolic Number

A Dissertation submitted in partial satisfaction of the requirements
for the degree Doctor of Philosophy

in

Experimental Psychology

by

Sebastian Holt

Committee in charge:

Professor David Barner, Chair
Professor Jonathan Cohen
Professor Judith Fan
Professor Victor Ferreira
Professor Federico Rossano

2025

Copyright

Sebastian Holt, 2025

All rights reserved.

The Dissertation of Sebastian Holt is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2025

TABLE OF CONTENTS

DISSERTATION APPROVAL PAGE iii
TABLE OF CONTENTS iv
LIST OF FIGURES v
LIST OF TABLES vi
LIST OF ABBREVIATIONS vii
ACKNOWLEDGEMENTS viii
VITA ix
ABSTRACT OF THE DISSERTATION x

INTRODUCTION 1
Perceptual Limits 2
How Symbolic Number Systems Overcome Perceptual Limits 3
How Children Learn Number Systems 7
REFERENCES 15

Chapter 1 Creating ad hoc graphical representations of number 18
Experiment 1 26
Experiment 2 36
Experiment 3 43
General Discussion 51
References 57
Supplemental Materials 61

Chapter 2 Learning a novel number system: The role of compositional rules & counting procedures 66
Experiment 1 79
Experiment 2 93
General Discussion 103
References 111
Supplementary Materials 117

Chapter 3 Can preschoolers learn the syntax of number? Using rules to combine familiar and novel number words 118
Experiment 1 126
Experiment 2 136
General Discussion 142
References 148

GENERAL DISCUSSION 152
REFERENCES 159

LIST OF FIGURES

Figure 1.1: 1-to-1 Representations of Quantity 22

Figure 1.2: Experiment 1 Diagram, Accuracy, and Strokes per Sketch 32

Figure 1.3: Some Example Sketches 34

Figure 1.4: Experiment 2 Diagram and Task Performance 39

Figure 1.5: Experiment 3 Diagram and Symbol Use 45

Figure 2.1: Experiment Phases 83

Figure 2.2: Counting in Recall and Generalization 86

Figure 2.3: Use of Different Rule-Types 86

Figure 2.4: Counting and Rules in Number-Word Learning 97

Figure 3.1: Trial Structure of Experiments 1 and 2 130

Figure 3.2: Accuracy for Small Sets, Large Sets, and Multiplier Tasks 135

Figure 3.3: Accuracy for Monkey and Giraffe Numbers 139

Supplemental Figure S1.2: VGG-19 Classifier Performance 64

Supplemental Figure S1.3: Spatial Clustering 65

LIST OF TABLES

Table 2.1: Several Base Systems 69

Table 2.2: Sample Base-3 Number System 81

Table 3.1: Consistent Strategies in Experiment 1 133

Table 3.2: Consistent Strategies in Experiment 2 142

LIST OF ABBREVIATIONS

CP	Cardinal Principle
PI	Parallel Individuation
ANS	Approximate Number System

ACKNOWLEDGEMENTS

I owe more thanks than I can ever convey to many friends and family who waited days, weeks, or months for me to stop thinking and write a message already. Thank you also to my committee for your helpful feedback on this and other work, and especially to Dave and Judy for your guidance, support, and patience over the years. I am grateful for the critical eye with which you approach the world and interact with me, and the compassion and tact with which you have responded to all my clumsiest efforts to do the same. To avoid delay in sending this out, I won't perseverate on choosing more words or naming names here – a drop in the ocean, at best – save to act on the good advice that you've often given me to not overthink it, and to say thank you for humoring me time and again with compelling reasons why.

Chapter 1, in full, is a reprint of the material as it appears in Holt, S., Fan, J. E., & Barner, D. (2024). Creating ad hoc graphical representations of number. *Cognition, 242*, 105665. The dissertation author was the primary researcher and author of this paper.

Chapter 2, in full, has been submitted for publication and may appear as Holt, S. & Barner, D. (2025). Learning a novel number system: The role of compositional rules & counting procedures. Under review at *Cognitive Science*. The dissertation author was the primary researcher and author of this paper.

Chapter 3, in full, has been submitted for publication and may appear as Holt, S. & Barner, D. (2025). Children's path to number syntax: concrete logic or procedural abstraction. Under review at *Philosophical Transactions of the Royal Society B.* The dissertation author was the primary researcher and author of this paper.

VITA

2019 Bachelor of Arts in Linguistics, Princeton University

2025 Doctor of Philosophy in Experimental Psychology, University of California San Diego

ABSTRACT OF THE DISSERTATION

Artificial Languages and Cognitive Constraints on Symbolic Number

by

Sebastian Holt

Doctor of Philosophy in Experimental Psychology

University of California San Diego, 2025

Professor David Barner, Chair

Research on the development of number concepts has tended to focus on existing systems of symbolic number (e.g. Arabic numerals), which are a small subset of the number systems attested in the historical record. Yet numbers also seem to exist independently of how we represent them in symbols, raising two questions. One is whether the study of numerical cognition is "overfit" to presently common forms of symbolic number. It may be that empirical regularities in the way that children learn number concepts, for example, owe more to

circumstantial features of the symbols that represent number than to perennial cognitive constraints on humans' ability to think about number. For example, children's ability to recognize arithmetical operations in the syntax of number words may depend on first learning how to count, but may be able to emerge independently of counting as well, through other forms of exposure to number words. A related question is how cognitive constraints have shaped the number systems that people do still use, e.g. by making them easier to learn or communicate. For example, is a base-10 system easier to learn than base-4? In three studies, participants' inferences about artificial number symbols are shown to be closely constrained by formal properties of the symbolic medium they learn. Chapter 1 asks how numerate adults used a graphical medium to communicate numerical information, when forbidden to use existing symbols. Chapter 2 asks how adults learn number words differently when exposed to number systems that vary in their reliance on syntactic composition or count routines. Chapter 3 asks how children interpret the compositional structure of number words, when those compositions are conjunctions of small number words, scaffolded by familiar linguistic constructions. Throughout, multiplicative rules pose more of a challenge for learners than cumulative representations do.

INTRODUCTION

Symbolic representations of large, exact number, like *thirty-three* or *five-million-and-eight*, have not always existed in human culture, and children learn number symbolic numbers through a protracted, years-long process that occurs in several stages. Yet, once we have learned number symbols, many properties of the number concepts they denote are immediately accessible from our symbolic representations of them. For example, it is easy to determine at a glance that 1000000 denotes a larger number than 100, but harder to judge that the Roman numeral MM denotes a larger number than MDCCLXXVII. How much of our intuition about number depends on arithmetic properties of the numbers themselves, and how much depends on features of the symbols we use to represent numbers, which could have been represented differently? The symbolic number systems that most of the world uses today are a small subset of the systems that have historically existed in human culture, and an even smaller subset of the systems that are conceivable. Yet, psychological theories of numerical cognition are mostly tested in populations who learn number systems with a limited range of characteristics. For example, almost all modern languages feature a number system with base-10 or base-20 rules of composition. Despite the considerable amount of exposure that people in numerate cultures receive to certain number systems, the empirical findings below suggest that their numerical reasoning can still be heavily influenced by the formal features of novel symbolic tools, and that numerical cognition research could gain much from studying how people interact with historically attested or entirely artificial symbolic number. This thesis presents three studies that examine how different kinds of number systems are created, learned, or reproduced among numerate adults (Chapters 1 and 2) and semi-numerate children (Chapter 3).

Perceptual Limits

Humans share with other animals several innate abilities to represent quantity, with different limitations. One of those abilities, supported by an approximate magnitude system (or approximate number system; the **ANS**), is the capacity to discriminate approximate quantities of arbitrary scale, an ability that becomes unreliable as magnitudes increase. For example, adult humans are easily able to discriminate magnitudes of 8:16, but not 15:16 (Dehaene, 1997; Whalen, Gallistel, & Gelman, 1999). This capacity is evolutionarily ancient, and even fish can complete a similar task with an acuity that surpasses many terrestrial vertebrates (Agrillo & Bisazza, 2017). Infants' ability to discriminate approximate magnitudes is relatively weak, but increases in acuity with age (Xu & Spelke, 2000; Halberda & Feigenson, 2008), and has even been argued to underlie the acquisition of the natural numbers (Gallistel & Gelman, 2000; Nieder, 2017; though see Carey & Barner, 2019, for a rebuttal of this claim). Another numerical ability is the ability to subitize small quantities. Subitizing means to perceive immediately the cardinality of (i.e. the number of entities in) a set, and is limited in humans to sets of about 3 or 4 objects at most (Jevons, 1870; Feigenson, Dehaene, & Spelke, 2004). Subitizing may be supported by parallel individuation (**PI**), the capacity to keep track of several separate indices at once (Piazza et al., 2011). Alternatively, it may be supported by the same mechanism as the ANS, but in the range of small numbers (Cheyette & Piantadosi, 2020). Either way, the system is capable of generating summary representations of the cardinality of a collection, such as what •••, III, abc, and .: all have in common. In such a representation, the property of •••-ness is perceived as a feature the entire collection of things as a whole, but is not a property of any single member taken alone. Because the representation is intuitively accessible, it can be named and manipulated symbolically. Humans are thus able to represent large, approximate quantities

supported by the ANS and small, exact quantities by subitizing. Large, exact number representations are beyond these perceptual limits, and require some culturally transmitted symbol system for people to learn and reason with.

How Symbolic Number Systems Overcome Perceptual Limits

Three mechanisms by which human cultures have overcome the above perceptual limits to represent large, exact quantities are tallies, count routines, and *ad hoc* comparisons. Tallies are a mode of ensuring 1-to-1 correspondence between two sets, such as placing a rock next to a gate every time a sheep leaves in order to keep track of how many sheep must be brought back at night. When the sheep return home, any rock from the pile can be removed to keep track of any one sheep. In this way, tallies take advantage of Hume's principle, that two sets are equal in number if and only if they are in 1-to-1 correspondence. Because of this, tallies rely on an enumeration procedure for their meaning. Formally, this is achieved by selecting one element at a time from the set of things that are being quantified and one element from the items in the tally, and then bracketing those items aside and repeating the procedure on the rest of each set.

A similar technology is the count system, which is like a tally in that its use entails enumerating two sets, but different in that each element of the count system is non-fungible with the others. Formally, this means that elements must always be selected from the set of tally symbols in exactly the same order. For example, body counting procedures are a type of count system, in that they designate a series of points on the body to serve as the measuring set. Within a culture, one point on the body is always selected first (say, the little finger), along with one element of the set that must be quantified. Then another point on the body is always selected second (e.g., the ring finger), along with another item in the set to be quantified, etc, until that set is exhausted. Most frequently, these systems begin by cycling through the fingers, but do so in a

wide variety of ways (Bender & Beller, 2012; Saxe, 2014). Whatever the diversity of tallies and counts, they rely on an enumerative procedure to derive their meaning.

On the other hand, ad hoc comparisons can be made without having to enumerate all the elements. Instead, they simply map a label directly onto a perceptually available quantity meaning, such as many Nadahup languages' words for 'three', which derive from the rubber tree seed pod, which comes in threes (Epps, 2006). These words often map directly to subitizable quantities, need not rely on 1-to-1 correspondence to understand, and in fact need not have precise cardinal meanings at all. For example, marriage practices among the Nadahup require that spouses get married concurrently with a sibling of the opposite sex. As a result, the exact number of potential spouses between two families is not of direct relevance, but whether that number is odd or even is relevant. This has left its mark clearly in the verbal number systems of those languages, such as Dâw, which uses "has a sibling" or "has no sibling" to refer to all even and odd quantities from 4 to 10. When a cardinal meaning must be conveyed, a tally is displayed on the fingers of the speaker to supplement the ad hoc comparator (Epps, 2006). Critically, and unlike tallies and body counts, they do not need to derive their conceptual content from a counting procedure, but can usually be subitized easily.

From their origins in tallies and ad hoc comparisons, number systems often evolve a generative syntax that facilitates representations of larger and more abstract quantities, such as 20 million. In order to express an unbounded number of meanings, a generative syntax must be combined with a compositional semantics. A generative syntax refers to a formal system composed of symbols – usually in a finite lexicon – that can be combined in infinitely many ways. A compositional semantics grounds such a syntax to meaning: by attaching meaning to every lexical item in the book of symbols, computing the syntactic operations allows the

meanings to also be combined, creating novel thoughts (famously, "Colorless green ideas sleep furiously" is conceivable because a generative syntax can compose ideas together from a semantics; Chomsky, 1957). In number systems, this work is done by a numerical base. For example, the vast majority of the world's current languages have a base-10 number system. These systems include 10 unit words (e.g., one, two, three… ten), beyond which the word for 'ten' is in some way combined with other words from the unit list to compose representations for larger numbers (e.g. 'seven-teen' as 7 + 10, or 'thir-ty-five' as 3*10 + 5). However, there are many other base systems as well (Hammarström, 2010; Comrie, 2013), such as the Chumash base-4 (Beeler, 1964), Mangarevan mixed base-10 and base-2 (Bender & Beller, 2014), or the base-6 use in many Papuan languages for counting yams (Comrie, 2013).

How does a finite set of number words become a generative number system with a compositional numerical base? Both the historical trajectory of quantifiers, and the experimental study of semiotic systems more generally, suggest that humans are capable of turning any finite symbolic system into a systematic, generative one. Cooperrider and Gentner (2019) trace the historical trajectory of many measurement terms that start out as ad hoc comparisons, but then undergo a process of conventionalization (e.g., everyone recognizes that a barleycorn can be used to represent a small length), standardization (e.g., when there is only one correct 'barleycorn length'), and finally systematization (e.g., defining the barleycorn as ⅓ of an inch). This process is often accompanied by increasing grammaticalization and attrition of morphological features, such as the history of 'football field' comparisons. Earlier in the 20th century, it became conventional to describe something as "*long as a football field*". Once the convention had been in place for long enough, people began to use a more succinct measure-word construction, such as "*four football fields long*" (Morris, 2017; Cooperrider & Gentner,

2019). This example illustrates a point that extends to numeral systems as well: grammaticalization changes the meaning of an ad hoc comparison into something that fits into a system, which is also precisely how many number words originated historically. Such is the case, for example, with the English word *five*, which derives ultimately from the Proto-Indo-European *pénk^{w}e, meaning *fist* or *hand* (Horowitz, 1992).

Semiotic systems that evolve in tightly controlled iterated learning studies show a similar pattern of systematization. In such studies, participants must learn a set of symbols that refer to a set of referents (e.g., learning that a word like *pilu* refers to a red square, and *gola* refers to a blue triangle). Critically, participants in iterated learning experiments reproduce the set of symbols from memory once they have learned it, typically producing some errors in the process. That reproduction is then given to a new participant to learn and reproduce, and the whole process is repeated over several generations (or back and forth between conversational partners or groups). The small errors that individuals make usually reflect inductive biases that extend perceived grammatical rules in the language, and tend to accumulate and make meaningful changes in the symbolic system. One of the most consistent findings across studies of artificial languages is that combinatorial syntax can emerge in a few generations (Kirby, Cornish, & Smith, 2008; Cornish et al., 2017; Verhoef, Kirby, & de Boer, 2014), or even within a single generation of communication partners, given enough time (Raviv, Meyer, Lev-Ari, 2019). As with the trajectory of measure words described above, the resemblance of the symbols to their referents may be lost when a symbolic system becomes more combinatorial (Hofer & Levy, 2019). However, this feature often has an important communicative function, with consequences for how easily the number system can be learned (Fay, Garrod, and Roberts, 2008).

How Children Learn Number Systems

Most developmental inquiry into the ontogeny of number concepts focuses on children acquiring the generative number systems of their first language(s). In such cultures (especially in the US) children typically begin by learning a rote count list, beginning at around 24 months of age. While they are able to recite many words of the count list, they do not recognize that number words are used to refer to quantity (Slusser & Sarnecka, 2011). Some months later, children come to recognize the precise meaning of the word for *1*, as measured by asking them to give one item to an experimenter (Wynn, 1992; Sarnecka & Lee, 2009). Several months later, those same children are able to give sets of one and two when asked to do so, but fail to reliably give larger amounts. After another few months, they also learn the meaning of words for *3*, and then *4*. During these stages, they are known as subset-knowers (or *N*-knowers, where *N* is the highest number word they are capable of relating to the conventional meaning). Some time after acquiring the number meaning of *three* or *four*, children make a large conceptual leap to become cardinal-principle knowers (CP-knowers). The cardinal principle (**CP**) means recognizing "that the last number used when accurately counting a set denotes the exact cardinal value of the set" (Gelman & Gallistel, 1978; Carey & Barner, 2019). At that point, children are able to provide the appropriate number of objects when asked for sets of any cardinality in their count list. For large sets, they typically do this by counting the objects, stopping at the number that was asked for, and giving the accumulated set.

A major question in the field concerns how it is that children make this leap from understanding item-based properties of a small, finite set of numbers to understanding systematic properties of the larger, open-ended range of numbers suggested to them by their knowledge of the count list. One proposal for how this happens is that a bootstrapping process occurs after

children learn 3-4 number words, whereby they generalize the relationships between the number concepts they already know to number concepts denoted by larger number words (Carey, 2004; Gentner, 2010). For example, according to Carey (2004), children are first assumed to represent small sets via parallel individuation. After gradually learning the meanings of the words 'one', 'two', and 'three', they notice that for every difference of one object between sets of one and two, or two and three, the number word labels of each set occur in adjacent positions of the count list (e.g., that *two* represents a set with 1 more item than *one,* and *three* represents one more than *two*). This observation in turn licenses the inference that the count list can be used to represent an ordered sequence of exact quantities that differ by 1, and that the count list can be used to generate sets of any cardinality – a hallmark of the cardinal principle (Carey, 2009). Under this account, the conceptual content required to make the leap to CP knowledge is only small-set representations, a finite count list, and the observation that an increment of one object in a physical set corresponds to an increment of one word in the count list. This last observation has an important property that is especially important: because it must generalize to all items in the count list (regardless of the size of the count list), it must have a recursive form. For this reason, it can be formalized as the successor function: every number *n* in their count list has a successor *n+1* that is larger by exactly one. If the successor function is necessary for CP-knowledge, then one should expect that all CP-knowers understand it.

Against this account, it is also possible that acquiring the cardinal principle is not due to an induction from representations of small sets to the rest of the count list. In fact, even children who are classified as CP-knowers often do not recognize that successive numbers in their count list differ by 1, which under bootstrapping theory is supposed to be the observation that licenses an induction in the first place (Davidson, Eng, & Barner, 2012). Other discoveries that the

bootstrapping account suggests should come with CP-knowledge also do not appear to materialize until later, such as the successor function (Cheung, Rubenson, & Barner, 2017), or that numbers go on forever (Chu et al., 2020). All of this evidence points away from a complete induction over the count list, and towards a set of rules that remain limited in their scope, again raising the question not only of what underlies CP-knowledge, but where later-emerging capacities come from, such as the successor function.

An alternative proposal to the bootstrapping account posits that discoveries like the successor function or infinity require children not to make inferences from their counting procedure, but to observe regularities in the syntactic rules that compose the number words (Cheung, Rubenson, & Barner, 2017). While the proximate source of the conceptual innovation under this account is number syntax rather than a count routine, mastery of the rote counting system may play a unique role in children's ability to learn from the combinatorial rules of number syntax. This is partly because it structures children's exposure to number words in way that makes analogies between syntactic and arithmetic properties more salient, e.g. successive number words are often minimal pairs. Experience with counting large numbers provides exposure to the rules of composition (e.g., the more often one counts to 100, the more often they observe that double digit numbers ending in *-nine* are succeeded by novel decade labels). One implication of this is that the more frequently a rule is observed, the more likely it is to be learned productively. Because of this, children learning number systems that clearly display the rules of composition should acquire concepts like the successor function faster than children learning more irregular number systems (Miller & Stigler, 1987). For example, some languages contain number words that correspond closely to the generative rules of the system. This is the case in Cantonese, where the number word meaning *12* (*sahpyi*) is composed directly of the

words meaning *10* (*sahp*) and *2* (*yih*). Other languages, such as Hindi, contain many irregular number words that have to be memorized independently of compositional rules. The standard Hindi expression of *12* (*baarah*) is not transparently composed of *10* (*das*) and *2* (*do*). In principle, the successor function should be easier to learn in Cantonese than in Hindi. Schneider et al. (2020) investigated the influence of these language differences on young children's understanding of the successor principle, and also carefully controlled for cultural variability by testing children from the same school in each of their three languages of instruction – English, Hindi, and Gujarati – which varied in morphosyntactic regularity. In one task, they first measured children's ability to count as high as they could. If children stopped because they did not know how to go continue, the experimenters then offered the name of the next number in the count list, and recorded how high children could count after being given one piece of help. The measure they took from this task was the difference between children's initial highest count and their final highest count (after being offered a missing number word), which they interpreted as a measure of children's productivity with the number syntax. For example, if a child could not continue counting even after being offered the next number word, they were not able to use the syntax productively. If a child stopped at, e.g., *thirty-nine*, were offered the word *forty*, and continued to count up to *fifty-nine*, then this was taken as an indication that they had learned at least part of the compositional rules of number syntax. With this measure, the authors found that children learning more predictable languages tended to be more productive users of their number system's syntax, even when students learning different languages attended the same school.

While the counting routine probably supports much of children's later number knowledge, it is a complex cultural artifact that not only exposes learners to many instances of syntactic rules, but also organizes number words into a unique order. How important is the

organization of number words into a sequence (e.g., “one, two, three…) for children to learn the compositional rules that underlie number words? It may be the case that compositional semantics in ordinary language use, e.g. in expressions such as “dogs and cats”, has the capacity to support the arithmetical rules of composition that children seem to acquire from counting (Spelke, 2017). By implication, perhaps rote counting is not strictly necessary for learning to compose number concepts at all. After all, children already use a compositional semantics every day by as part of language. Other features of natural language have already been used to explain some cross-linguistic differences in the acquisition of the earliest number words. For example, while English has two grammatical numbers – singular and plural – Slovenian has three – singular, dual, and plural. Because the grammar of their native language marks sets of two differently from other sets, Slovenian children are faster than American children to acquire the number word meaning *2* (Almoammer et al., 2013). Furthermore, some of the world’s languages have precise, compositional number words that are never part of a count routine. The Western Tribe of Torres Straits, for example, had a set of number words the entirety of which is (1) urapun, (2) okosa, (3) okosa urapun, (4) okosa okosa, (5) okosa okosa urapun, (6) okosa okosa okosa (Haddon, 1890). This system was limited in range, but it did contain a compositional semantics. Critically, each of the two lexical primitives and some of their compositions can be easily mapped directly to a perceptual representation of small sets (sets of one, two, and three). In principle, this system had generative capacity, and it did not require a counting routine to master or use generatively.

Prior studies testing children’s ability to reason compositionally about number words have found that (1) children learning more syntactically regular number systems (i.e. fewer exceptions to rules) are be better able to parse syntactically complex number words and (2) CP knowers can interpret a made-up expression for 2 groups of 3 regardless of their first language.

Guerrero et al. (2020) asked whether children could spontaneously decompose complex numbers by presenting children with blocks arranged in groups of 10, and then asking them to give large sets of blocks, for example 32. They measured whether children would recognize that 32 is the composition of *3*10* and *2*, and give three groups of ten blocks and two remainder. English did not learn to do this until quite late, and although Korean speaking children (a language with a very regular number system) completed the task using grouped 10s earlier, children in either language who completed the task successfully were able to count quite high, past 99 (Korean) or 109 (English). As with other studies however, the number systems being investigated do not show any kind of composition until after the number 10. One recent study by Cheung, Dale, and le Corre (2016) did investigate children's understanding of smaller compositional numbers by teaching them an artificial number word. They asked whether 4- to 6-year-old English and Cantonese speaking children could learn a novel numerical construction that referred to groups of three objects (e.g. "There are one *gobi* books" to mean three houses), and generalize the construction to other multiples of three (e.g. "Who has two *gobi* books?"). This novel construction paralleled the syntactic structure of constructions like *two hundred*, which are composed of a digit, *two*, and a multiplier, *hundred*. Such constructions are syntactically analogous to noun phrases composed of a determiner and a noun (e.g., *a dog*), but do not appear transparently in English number word morphology until 200, or until 20 in Cantonese - both of which are far beyond the subitizing limit & number beyond which children are typically labelled CP-knowers. Both English and Cantonese speaking children as young as 4 were able to generalize the construction, provided that they had learned the novel number word as a multiplier within a digit-multiplier construction (i.e. "one *gobi* books" but not "*gobi* books"). Other, theoretical approaches towards symbolic system learning have also focused on system learning

per se, rather than the acquisition of single concepts or rules. Rule, Dechter, and Tenenbaum (2016) point out that number concepts are necessarily part of systems, and that a model sensitive to compositional rules shows a similar developmental trajectory to children learning to count, and produces similar errors similar to those that children produce when counting. Although they apply it to learning counting rules, it could in principle be applied to other types of input. Other attempts to teach children novel counting systems, using animals (Jacobs, Flowers, & Jara-Ettinger, 2021) or alphabet letters (Schneider, unpublished data) as count indices, have not involved combinatorial syntax. Jacobs, Flowers, and Jara-Ettinger's (2021) success in teaching CP-knowers a novel counting system had two particularly interesting outcomes, however. One is that children were at chance when reasoning about quantities denoted by animal names, if those animal names are used as natural language quantifiers (i.e. "There are frog blocks in this box"), but were successful when the order of the animals was relevant because of an interpretable context (i.e. feeding each animal in a particular order, and needing to know which animals could be fed with the available food). This may be because animal names already have a meaning in children's mental lexicons, and cannot be flexibly used in another syntactic context with an unrelated meaning. The second outcome is that even if animal names were used in the novel syntactic context as a quantifier, success on the task was facilitated if the size of the animals was congruent with their order in the count list (and, by extension, the cardinality they represented). This was true only if larger animals represented larger quantities.

The goal of this dissertation is to show how artificial language methods can be used to shed light on the mutual influences between formal properties of symbolic number systems on one hand, and learners' own cognitive and communicative constraints on the other. Three very different empirical studies are presented, testing both adult and child participants in a numerate

culture, as they learn to use artificial number systems. In Chapter 1, dyads coordinated about numerical information using a graphical communication medium. Chapters 2 and 3 ask how number systems that are compositional in a much smaller range than typically studied might change the kinds of concepts that learners are able to use in those systems. In Chapter 2, adults learned and generalized the syntax of number systems with different numerical bases (Experiment 1), and number systems that they learned via a counting procedure or by rehearsing number words in no particular order (Experiment 2). In Chapter 3, an experiment with children is reported, asking how they understand and generalize the compositional rules of a novel number system, when these rules of composition are explicitly expressed by conjoining two familiar English number words together with the familiar word *and*.

REFERENCES

Agrillo, C., Petrazzini, M. E. M., & Bisazza, A. (2017). Numerical abilities in fish: a methodological review. *Behavioural processes*, *141*, 161-171.

Almoammer, A., Sullivan, J., Donlan, C., Marušič, F., Žaucer, R., O'Donnell, T., & Barner, D. (2013). Grammatical morphology as a source of early number word meanings. *Proceedings of the National Academy of Sciences*, *110*(46), 18448-18453.

Beeler, M. S. (1964). Ventureño numerals. *Studies in Californian linguistics*, 13-18.

Bender, A., & Beller, S. (2012). Nature and culture of finger counting: Diversity and representational effects of an embodied cognitive tool. *Cognition*, *124*(2), 156-182.

Bender, A., & Beller, S. (2014). Mangarevan invention of binary steps for easier calculation. *Proceedings of the National Academy of Sciences*, *111*(4), 1322-1327.

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*.

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.

Cheyette, S. J., & Piantadosi, S. T. (2020). A unified account of numerosity perception. *Nature human behaviour*, *4*(12), 1265-1272.

Chomsky, N. (1957). Syntactic structures. In *Syntactic Structures*. De Gruyter Mouton.

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*.

Cooperrider, K., & Gentner, D. (2019). The career of measurement. *Cognition*, *191*, 103942.

Cornish, H., Dale, R., Kirby, S., & Christiansen, M. H. (2017). Sequence memory constraints give rise to language-like structure through iterated learning. *PloS one*, *12*(1), e0168532.

Davidson, K., Eng, K., & Barner, D. (2012). Does learning to count involve a semantic induction?. *Cognition*, *123*(1), 162-173.

Dehaene, S. (1997). The Number Sense: How the mind creates mathematics London: Allen Lane.

Epps, P. (2006). Growing a numeral system: The historical development of numerals in an Amazonian language family. *Diachronica*, *23*(2), 259-288.

Fay, N., Garrod, S., & Roberts, L. (2008). The fitness and functionality of culturally evolved communication systems. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1509), 3553-3561.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in cognitive sciences*, *8*(7), 307-314.

Gallistel, C. R., & Gelman, R. (2000). Non-verbal numerical cognition: From reals to integers. *Trends in cognitive sciences*, *4*(2), 59-65.

Gelman, R., & Gallistel, C. R. (1978). The child's concept of number. *Cambridge, MA: Harvard*.

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

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*, *19*, 297-440.

Halberda, J., & Feigenson, L. (2008). Developmental change in the acuity of the" Number Sense": The Approximate Number System in 3-, 4-, 5-, and 6-year-olds and adults. *Developmental psychology*, *44*(5), 1457.

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

Hofer, M., & Levy, R. P. (2019). Iconicity and structure in the emergence of combinatoriality.

Horowitz, F. E. (1992). On the Proto-Indo-European etymon for 'hand'. *Word*, *43*(3), 411-419.

Jacobs, C., Flowers, M., & Jara-Ettinger, J. (2021). Children's understanding of the abstract logic of counting. *Cognition*, *214*, 104790.

Jevons, W. S. (1871). The power of numerical discrimination. *Nature*, *3*(67), 281-282.

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences*, *105*(31), 10681-10686.

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

Morris, C. (2017). The size of things: An ngram experiment [Blog post].

Nieder, A. (2017). Number faculty is rooted in our biological heritage. *Trends in cognitive sciences*, *21*(6), 403-404.

Piazza, M., Fumarola, A., Chinello, A., & Melcher, D. (2011). Subitizing reflects visuo-spatial object individuation capacity. *Cognition*, *121*(1), 147-153.

Raviv, L., Meyer, A., & Lev-Ari, S. (2019). Compositional structure can emerge without generational transmission. *Cognition*, *182*, 151-164.

Rule, J., Dechter, E., & Tenenbaum, J. B. (2015). Representing and Learning a Large System of Number Concepts with Latent Predicate Networks. In *CogSci* (pp. 2051-2056).

Sarnecka, B. W., & Lee, M. D. (2009). Levels of number knowledge during early childhood. *Journal of experimental child psychology*, *103*(3), 325-337.

Saxe, G. B. (2015). *Culture and cognitive development: Studies in mathematical understanding*. Psychology Press.

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.

Slusser, E. B., & Sarnecka, B. W. (2011). Find the picture of eight turtles: A link between children's counting and their knowledge of number word semantics. *Journal of experimental child psychology*, *110*(1), 38-51.

Spelke, E. S. (2017). Core knowledge, language, and number. *Language Learning and Development, 13*(2), 147-170.

Verhoef, T., Kirby, S., & De Boer, B. (2014). Emergence of combinatorial structure and economy through iterated learning with continuous acoustic signals. *Journal of Phonetics*, *43*, 57-68.

Whalen, J., Gallistel, C. R., & Gelman, R. (1999). Nonverbal counting in humans: The psychophysics of number representation. *Psychological science*, *10*(2), 130-137.

Wynn, K. (1992). Children's acquisition of the number words and the counting system. *Cognitive psychology*, *24*(2), 220-251.

Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition*, *74*(1), B1-B11.

Chapter 1

Creating *ad hoc* graphical representations of number

Sebastian Holt[1], Judith E. Fan[1,2], & David Barner[1,3]

[1] Department of Psychology, University of California, San Diego

[2] Department of Psychology, Stanford University

Abstract

The ability to communicate about exact number is critical to many modern human practices spanning science, industry, and politics. Although some early numeral systems used 1-to-1 correspondence (e.g., 'IIII' to represent 4), most systems provide compact representations via more arbitrary conventions (e.g., '7' and 'VII'). When people are unable to rely on conventional numerals, however, what strategies do they initially use to communicate number? Across three experiments, participants used pictures to communicate about visual arrays of objects containing 1-16 items, either by producing freehand drawings or combining sets of visual tokens. We analyzed how the pictures they produced varied as a function of communicative need (Experiment 1), spatial regularities in the arrays (Experiment 2), and visual properties of tokens (Experiment 3). In Experiment 1, we found that participants often expressed number in the form of 1-to-1 representations, but sometimes also exploited the configuration of sets. In Experiment 2, this strategy of using configural cues was exaggerated when sets were especially large, and when the cues were predictably correlated with number. Finally, in Experiment 3, participants readily adopted salient numerical features of objects (e.g., four-leaf clover) and generally combined them in a cumulative-additive manner. Taken together, these findings corroborate historical evidence that humans exploit correlates of number in the external environment – such as shape, configural cues, or 1-to-1 correspondence – as the basis for innovating more abstract number representations.

Keywords: numerals, drawing, communication, counting, symbols

—

Representations of large exact number - like the numerals *15* and *2028* - have fueled some of the most important achievements of human societies. By allowing humans to represent large quantities exactly, numerals have provided a critical foundation to the development of mathematics and science, the coordination of economic activity in large groups, and the development of human infrastructure and technology. However, humans have not always had the capacity to precisely encode numbers, and many humans currently remain either partially or fully innumerate (Gordon, 2004; Pica et al., 2004; Frank et al., 2008). Further, the anthropological record indicates that written numeral systems evolved slowly over a period of thousands of years, concurrent with changes in complex social institutions and concrete material environments (Høyrup, 1994; Robson, 2008; Malafouris, 2010; Schmandt-Besserat, 2010; Chrisomalis, 2010; Beller et al., 2018). Finally, experimental research in psychology indicates that, prior to acquiring verbal and written numerals, children are unable to represent exact quantities larger than 3-4, and that in absence of such symbols, perception of number is severely limited (Carey, 2004; Carey & Barner, 2019; Feigenson, Dehaene, & Spelke, 2004). These observations raise the question of how our ancestors converged upon numerals as a solution to representing exact number, and why such innovation was so elusive in the history of our species. What psychological mechanisms might explain the process by which humans create and share novel representations of large exact number?

In the absence of a numeral system, humans, like many other species, represent large quantities approximately, relying on the approximate magnitude system (or AMS). While the AMS is able to represent quantities up to 3 or 4 with precision, representations become noisier as quantities increase, and sets are discriminated on the basis of their ratio, compatible with Weber's law (Dehaene, 1997; Whalen, Gallistel, & Gelman, 1999). Evidence of the AMS is

found across human cultures, independent of their adoption of exact numerical symbols. First, when numerate adults are asked to discriminate two sets of rapidly presented dots, they readily identify the larger set when the ratio between the sets is large (e.g., 40 vs. 80 dots, or 2:1), but struggle as the ratio becomes smaller (e.g., 70 vs. 80 dots, or 7:8). Second, innumerate adults, such as the Mundurukú, an Amazonian indigenous group, easily discriminate dot arrays that stand in a 2:1 ratio (e.g., 20 dots vs. 10 dots), but show declining performance for tighter ratios, like 3:2 (Pica et al., 2004). Similar results have been found in studies of innumerate Pirahã adults. For example, when shown a set of 8 objects and asked to match that set with another set of objects, Pirahã participants often provide approximate matches, and are precise for only the smallest sets up to around 3-4 (Gordon, 2004; Everett & Madora, 2011; cf. Frank et al., 2008). Similar results are found in US children who have not yet learned to count (Schneider et al., 2022). Finally, evidence for the AMS is also found in preverbal human infants and in non-human animals, including birds, rodents, and fish (Brannon & Merritt, 2011; Dehaene, 1997).

Numeral systems transcend the limits of the AMS by providing symbols and operations that differentiate large quantities exactly. Verbal numerals, body count systems (Bender & Beller, 2011; Comrie, 2011; Epps, 2006; Hammarström, 2010; Saxe, 1981), written numerals (Changizi & Shimojo, 2005; Chrisomalis, 2020; Ifrah, 2000), and physical calculators like the abacus (Frank & Barner, 2008; Hatano & Osawa, 1983; Stigler, 1984), use similar strategies to extend the human ability to quantify large sets (for reviews, see Barner, 2017; O'Shaughnessy et al., 2021). Often, these systems use 1-to-1 correspondence to represent the smallest numbers - using 4 fingers, abacus beads, or vertical lines to represent sets of 4 things (see Figure 1.1A). For example, clay envelopes created as early as 11,000 years ago in Mesopotamia often used 1-to-1 correspondence to indicate the number of tokens they contained (Schmandt-Besserat, 2010).

Also, a 1-to-1 strategy is found in more recent written number systems such as the familiar Roman numerals (e.g., I, II, III), but also in ancient Greek, Hittite, Cretan, Aramaic, Mayan, and other systems (Ifrah, 2000). When such systems are extended to represent larger numbers, they generally do so in one of two ways. Some systems use configural strategies such as horizontal spacing (e.g. 'IIII II' to represent '6') or stroke directionality (e.g. '~~IIII~~ I'; Chrisomalis, 2020). Also, they generally exhibit similar rules, only allowing new chunks to be created when all previous chunks attain a maximum value (e.g., allowing III III II, but not IIII II II). Other systems however, use arbitrary conventions to express larger numbers. For example, the Roman numeral system represents 5 as V and 10 as X. Likewise, the Soroban abacus **(**Figure 1.1B**)** uses horizontal space to represent place value, such that some beads have a value of 1, others a value of 10, and others 100, and uses vertical space to assign some beads (i.e., "heavenly beads") values of 5, 50, 500, etc. (e.g., Frank & Barner, 2008; Hatano & Osawa, 1983; Stigler, 1984). Finally, count systems based on the human body typically use arbitrary positions to represent numbers beyond 20 (Bender & Beller, 2011; Saxe, 1981).

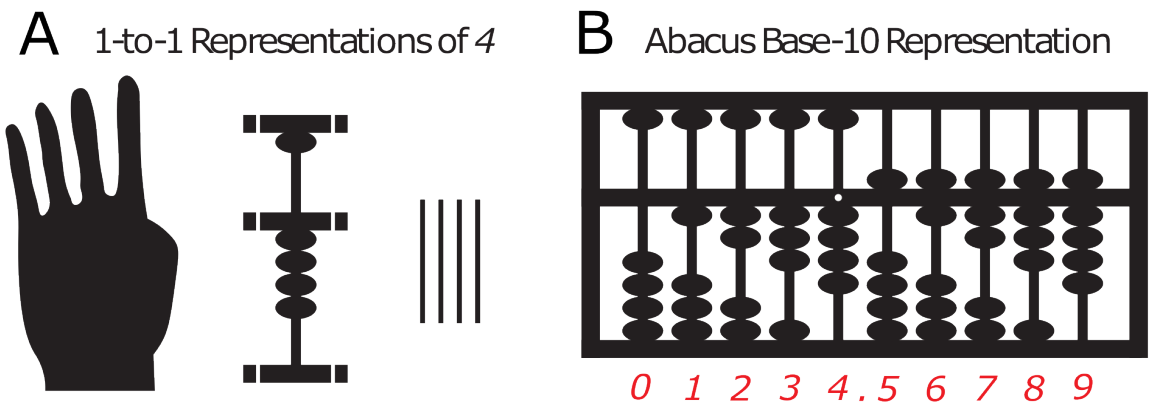


Figure 1.1. A. Three 1-to-1 modes of representing the quantity 4: fingers, abacus beads, and written strokes. **B.** A soroban showing different digits in base-10. Much like the western Arabic numerals, each column denotes a place-value, such that the column showing *3* represents a magnitude of *30* while the column showing *4* represents a magnitude *4*. The white dot indicates the ones-place, to the right of which columns represent decimal values.

Such examples raise the question of how arbitrary conventions arise that transcend 1-to-1 correspondence. According to one cultural-historical account, arbitrary conventions emerge from forms that are initially grounded in contextually available *ad hoc* comparisons to numerically correlated features of the environment (Cooperrider & Gentner, 2019). For example, measurement terms like "foot" are often created via comparisons to immediately available physical objects (e.g., one's own foot). These concrete measuring conventions are then shared with other people via acts of communication, which leads to standardization (e.g., defining a "foot" as the same length regardless of who uses it), and systematization, where they become defined by other units within a broader network of concepts rather than by their initial, non-arbitrary, *ad hoc* comparison (e.g., a foot is defined as 12 inches, ⅓ of a yard, etc, but *not* the foot of whoever is making the measurement). Number words often follow a similar trajectory, beginning via *ad hoc* comparison to concrete objects, but culminating in more arbitrary conventions over time as they become integrated to a new system of use. For example, in the Hup language (spoken in Colombia and Brazil) the word for 1 originated from the demonstrative term "that", 2 from "eye-quantity", and 3 from "rubber tree seed quantity" (since seed pods contain 3 seeds; Epps, 2006). Similarly, 10 is expressed as "both hands" and 20 as "both feet" - a widely attested practice in the cultural history of number (see also Bender & Beller, 2012; Saxe, 2014; Williams, 1940; O'Shaughnessy et al., 2021). However, as in the case of measure words, these expressions that began as *ad hoc* comparisons have subsequently changed, becoming phonologically distinct from the words that served as their historical basis, and used strictly for the abstract function of referring to number (and not to eyes, seeds, etc.). Such facts suggest that *ad hoc* reference is a critical first step in the process of creating number conventions, followed

by the derivation of abstract meanings from their position within the system of symbols (Damerow, 1996).

While the emergence of numerical conventions in cultural history is a topic of significant debate among anthropologists and historians (see Chrisomalis, 2004, for review), it is less frequently studied by cognitive psychologists, and we know of no systematic experimental study of how humans create novel numerals. Psychologists frequently discuss the evolutionary origins of numerical perception (Gallistel, Gelman, & Cordes, 2006), but few discuss the role of cognitive factors in the cultural evolution of numeral systems, with some notable exceptions (for some discussion of this topic, see Barner, 2017; Hurford, 1987; O'Shaughnessy et al., 2021; Xu et al., 2020). For example, in one recent study, Xu, Liu, and Regier (2020) analyzed the verbal numeral systems of 30 currently spoken languages, and argued that the evolution of cross-linguistic variation is shaped by a need for precise yet cognitively efficient communication. Similarly, O'Shaughnessy et al. investigated the cultural origins of symbolic number by characterizing the variability in attested numeral systems, and reviewing known differences in how different cultures and linguistic forms express large and small numbers. However, while experimental methods have sometimes been used to characterize variability in existing numeral systems (e.g., Bender & Beller, 2012; Frank et al., 2008; Gordon, 2004; Pica et al., 2004; Saxe, 2014; Gibson et al., 2019), no previous work has investigated the creation of novel numerals experimentally.

Although there is little experimental work related to the evolution of numerical symbols, a larger literature has used experimental methods to investigate the origin of symbols representing non-numerical concepts (Christiansen & Kirby, 2003; Kirby, Cornish, & Smith, 2008). Multiple studies have found that abstract conventional symbols emerge gradually in

interactive communicative games. In Pictionary-style games, while participants initially create drawings that are transparently related to the things they represent, later drawings tend to become more abstract and conventionalized (Garrod et al., 2007; Fay et al., 2010; Hawkins et al., 2023). An initial reliance on non-arbitrary symbols can also be found in novel communicative modalities, with which participants have no previous experience. For example, participants have been found to use more complex auditory signals to refer to more complex objects (Verhoef, Kirby, & De Boer, 2016; Hofer & Levy, 2019).

Here we adapt these experimental methods to investigate how participants use a graphical communication medium (i.e., involving mark-making and/or pictures) to convey number. In particular, we asked how participants express number when prevented from using pre-existing conventions (e.g., Arabic numerals), and how their ability to communicate is impacted by the availability of physical shapes and configurations that could be used as proxies for number, akin to the use of hands, feet, or words for objects like eyes and seeds. In Experiment 1, participants were paired up to play a drawing-based communication game in which one participant (the sketcher) sought to communicate about an array of objects to another participant who could not see the array (the viewer). In "number games," only the number of objects was relevant; in "shape games," only the identity of the objects was relevant. We found that drawings produced in the number games were markedly different from those in the shape games: they were generally composed of marks that stood in 1-to-1 correspondence with the objects but no longer resembled them - a form of abstraction that is typical in attested numeral systems. Moreover, some sketches seemed to mirror the spatial configuration of target sets, while others arranged the objects into orderly configurations. Following up on this observation, Experiment 2 further explored the conditions under which participants might exploit configural cues to express

number. When the sketcher was shown target arrays in the same configuration as the viewer, and was aware of this fact, sketchers frequently exploited these configural cues to use strategies other than 1-to-1 correspondence. While Experiment 2 provided an in-principle demonstration that participants used shared features of their physical environment to communicate number, it did so in a way that is likely rare in the wild, as most sets seldom occur in reliable configurations like rows and columns. Also, in both Experiments 1 and 2 we required participants to create representations of number *de novo*, whereas people often use existing objects and their labels to express number via *ad hoc* comparison to other objects. In keeping with these considerations, in Experiment 3 we asked participants to communicate number using a menu of familiar shapes that differed according to numerically salient features (e.g., a pair of cherries, a four leaf clover, etc.). We measured to what degree participants would seize upon these numerical features to transcend the use of 1-to-1 correspondence, and also what rules of composition they might use to express larger numbers.

Experiment 1

In Experiment 1, we conducted a preliminary investigation of how people use *ad hoc* graphical representations to communicate number (i.e., without the aid of existing symbols). Participants were paired up online to play a sketching-based reference game and, on each trial, were presented with four arrays of objects. One participant (sketcher) aimed to produce a sketch that would help their partner (viewer) identify which array they intended to refer to. To explore the impact of having the specific need to communicate about number information, we manipulated whether these arrays differed in the number or the kind of objects they contained. We sought to

characterize what strategies dyads used to communicate when only number information was relevant, by comparison to the baseline scenario in which only object identity was relevant. Towards this end, we both analyzed the visual properties of these sketches and recruited naive participants (recognizer) to interpret them out of context, providing complementary insight into their content and organization.

1.1 Methods

Experimental methods for this and subsequent experiments were pre-registered using the Open Science Framework, available at: https://osf.io/4q3t9/.

1.1.1 Participants

Participants were recruited for two tasks: a communication task and a recognition task. For the communication task, we initially recruited 134 participants from Amazon Mechanical Turk (AMT), who were paired up to form 67 dyads who interacted with one another throughout an entire experimental session. Data from six sessions were excluded according to our pre-registered criteria: four did not meet the performance threshold of 50% accuracy, while two others did not follow task instructions (e.g., having "drawn" text). Thus data from 122 participants (N=61 sessions) were included in further analysis. Of these participants, 117 individuals completed our optional demographic survey (53 female; age data was not collected due to a technical error). For the recognition task, we recruited a total of 211 participants from Amazon Mechanical Turk. Thirteen recognition participants were excluded according to pre-registered criterion of missing any one of four catch trials, and eight further participants were excluded for having participated in the communication task, leaving 190 participants, of whom

165 completed our optional demographic survey (62 female; M_{age} = 37.65 years, SD_{age} = 10.80 years). All participants provided informed consent in accordance with our IRB.

1.1.2 Stimuli

Stimuli were visual arrays of objects, each containing a variable number of identical shapes arranged in an arbitrary configuration. These arrays could contain between 1-8 items and contain one of 4 different shapes (i.e., bear, deer, owl, rabbit silhouettes), resulting in 32 number × shape combinations. This range of numbers was chosen to include values both above and below the subitizing range of ~3-4. To preclude reliance on spatial cues, the configuration of objects within each array was independently randomized for each participant and across dyads.

1.1.3 Procedure

Communication Task. The communication task was a sketching-based reference game for two players (Figure 1.2A), in which one participant (the *sketcher*) created sketches to help their partner (the *viewer*) identify one 'target' visual array at a time from three similar distractors. While the complete set of target arrays was identical across all games (i.e., every game featured sets of 1-8 bears, deer, owls, or rabbits), the distinguishing feature between target arrays and distractors within each trial varied between two conditions. Some dyads were randomly assigned to the Shape condition (N=29), in which all four visual arrays on a given trial featured the same number of animals but differed in the kind of animal they contained. Other dyads were assigned to the Number condition (N=32), in which all four visual arrays featured the same kind of animal but differed in the number of animals per array. For each array shown to the sketcher, a corresponding array with the same number and type of animal was shown to the viewer, but the

spatial configuration of animals was independently randomized for each participant, to preclude reliance on configural cues. Both participants in the dyad therefore had the same basis for inferring the which feature of the visual arrays – number or shape – was relevant.

On each trial, the sketcher used a 500×500px digital canvas embedded in a web browser to produce their sketch. They were required to complete this sketch in less than 30 seconds, after which additional strokes could not be registered. Also, they were told that they should not use existing numerical symbols (e.g., 5, 6, 7) but that they could create new symbols of their own (for complete instructions, see Supplemental Materials, section S1.1). After the sketch was complete, it was shown to the viewer, who was then asked to select the array they thought was the target, without a time limit. Then both participants received feedback: the viewer was shown the identity of the target and the sketcher was shown which array the viewer selected. There were 32 trials in each game, such that each of the 32 number × shape combinations served as the target exactly once. These 32 trials were divided into four blocks of eight trials each, such that each number appeared once within each block and each animal appeared twice.

Recognition Task. Participants in the recognition task were presented with sketches produced in the communication task, and were asked to identify shape or number information in each sketch. On each trial, these participants were presented with a single sketch and several buttons below. Buttons were either labeled with Arabic numerals or animal shapes, depending on whether the participant had been assigned to the Number or Shape recognition condition. Participants in the Number group judged which numeral lying in the range 1-8 best matched each sketch; participants in the Shape group judged which shape (i.e., bear, deer, owl, or rabbit) provided the best match.

Each participant was presented with a total of 61 sketches, one randomly sampled from each communication game (N=61 dyads). They were thus presented with sketches from both Number and Shape communication games, and had to guess intended meanings while naïve to the communicative context in which the sketches were originally produced. As an attention check, four catch trials were inserted at regular intervals throughout each session, where participants were presented with a sketch that matched one of the buttons exactly. Participants were included in analysis only if they chose the matching button in all four of these catch trials.

1.2 Results

1.2.1 Communicative Strategy. We first asked whether participants communicated number or shape information comparably. To this end, we measured (1) how accurate pairs were, (2) how long sketchers took, (3) how much virtual 'ink' sketchers used, and (4) how many strokes sketchers used. We constructed mixed-effects regression models to predict each of the above outcome variables, based on: communication condition (Number vs. Shape), trial block number (i.e., 1-4), shape (i.e., bear, deer, owl, rabbit), and the number of objects in the target array (i.e., 1-8). In all linear models, variation across games was modeled by fitting a random intercept for each game.

These models revealed similarly high accuracy in both conditions (b = -0.221, z = -0.623, p = 0.534; see Figure 1.2B), but that Shape game sketchers used more time and 'ink' to create their sketches (time: b = 6.288, t = 7.675, p < .001; ink: b = 0.034, t =8.98, p < .001). Over successive trial blocks, dyads in both conditions became more accurate (b = 0.494, z = 5.745, p < 0.001), took less time (b = -1.369, t = -23.051, p < .001), used less ink (b = -0.001, t = -4.742, p < .001), and made fewer strokes (b = -0.553, t = -11.029, p < .001). These results suggest that

communication of number and shape was functionally comparable, both in getting the point across and in doing so better over time.

We found that participants were more accurate for smaller cardinalities (b = -0.252, z = -6.001, $p < 0.001$), and produced more strokes for larger cardinalities in Number games (b = 1.029, t = 15.782, $p < .001$) but not Shape games (b = - 0.734, t = -7.767, $p < .001$), suggesting that a 1-to-1 strategy was used only for communicating number (Figure 1.2C). To explore this, we therefore analyzed the 'stroke ratio' of sketches. We reasoned that if the ratio of strokes (in a sketch) to objects (in a target) was 1, this would indicate the use of 1-to-1 correspondence (e.g., IIII to represent 4 objects), while a smaller number of strokes would indicate the use of a compressed form (e.g., IV). With this measure, we found that 77.2% of Number-game sketches used 1-to-1 correspondence (CI: [74.7, 79.8]) versus only 14.3% of Shape-game sketches (CI: [12.1, 16.6]).

A qualitative analysis of the sketches (see Figure 1.3) produced in these games revealed that Number-game sketches sometimes used strategies that we had not anticipated. One strategy featured the use of 1-to-1 correspondence, but was partially depictive, such that the configuration of objects within the target array was mirrored by the configuration of strokes within the sketch. Another strategy preserved 1-to-1 correspondence, but not configural information, instead rearranging strokes into patterns that appeared more geometrically regular than the random arrays of objects in the target. A rarer strategy appeared to make use of geometric shapes that correlated with the cardinality of the target array. For example, one participant used a five-pointed star to communicate an array of 5 owls. In general, such strategies are consistent with historical observations that humans draw on publicly available correlations in the environment to communicate about number (e.g., Epps, 2006).

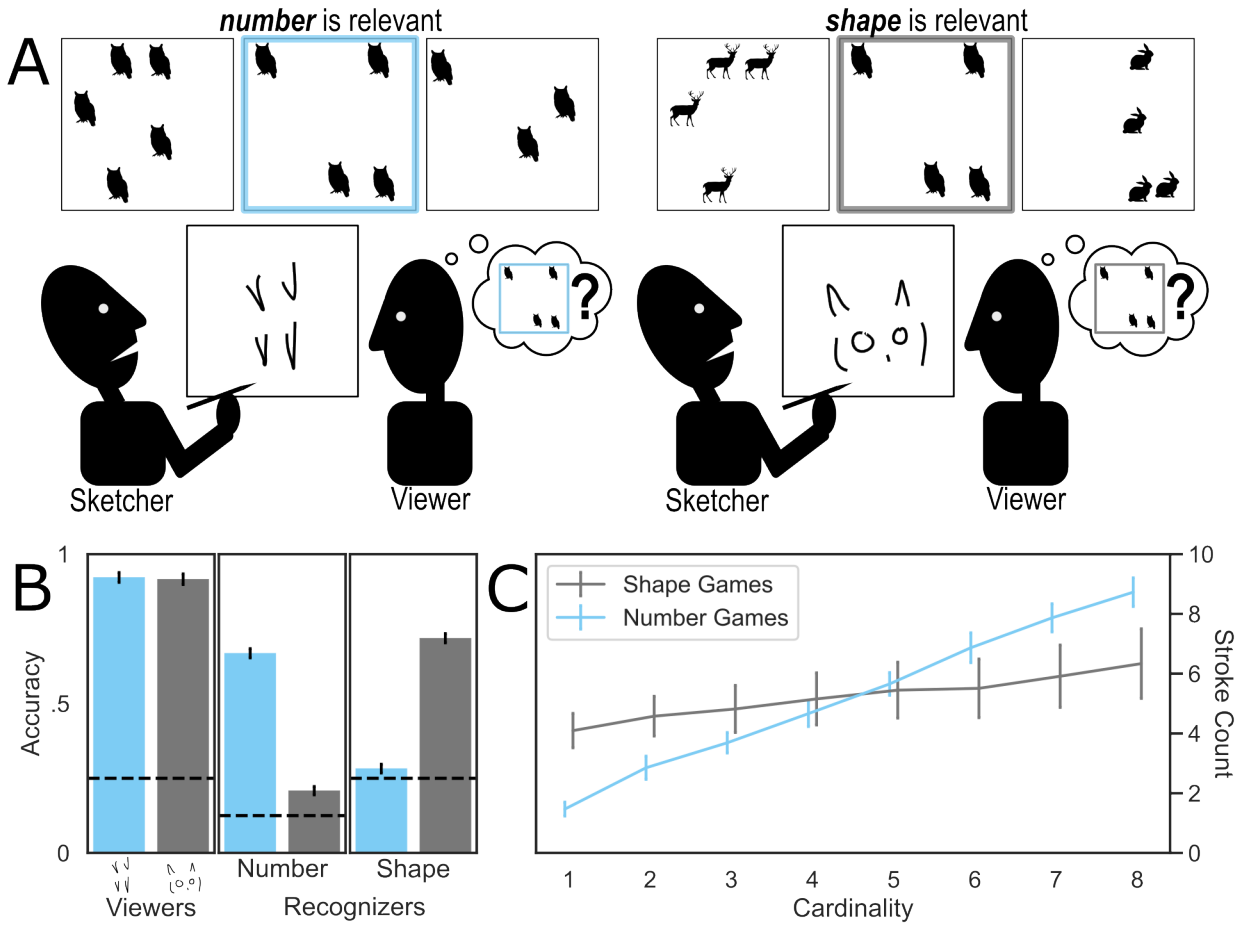


Figure 1.2. A. In the number condition (left), sketchers had to communicate one target image to a viewer from a set of images that differed in the number of objects. In the shape condition (right), the set of images differed in the kind of object, but not the number. B. Left panel: viewer accuracy for games in each communication task condition. Middle panel: Recognizer accuracy in providing number labels for sketches produced in the number condition (blue) and shape condition (gray). Chance performance in dotted lines. Right panel: same as middle, but for recognizers providing shape labels. C. Average number of strokes in sketches for each tested cardinality in both number (blue) and shape (gray) games. While a weak correlation exists in shape games, it is almost exactly 1 in number games (error bars are 95% CI in all panels).

1.2.2 Message Content. We next asked how much number or shape information was included in sketches, and whether this was affected by the relevance of each type of information (e.g., did Shape-game sketchers include irrelevant numerical information?). To answer this question, we asked each recognizer in the recognition task to recover one type of information – number or

shape – from sketches produced in the communication task. Each recognizer was presented with one sketch from every game (both Number and Shape games), and we measured the content of sketches by recognizers' ability to recover this information.

We constructed a binomial linear mixed-effects model, which included communication game condition (i.e., number or shape), recognition task condition (i.e., number or shape), communication game trial block number (i.e., 1-4), and the interaction between communication game condition and recognition task condition as fixed-effects predictors, and modeled variation between individual recognizers, communication games, stimulus cardinality, and shape using random intercepts.

Overall recognition accuracy was slightly lower for shape games than for number games (b = -2.364, z = -23.860, $p < .001$), and was also lower for recognizers trying to recover shape information rather than number information (b = -1.732, z = - 20.109, $p < .001$). Crucially, recognizers in both conditions were better at recovering information that was relevant in the original communication task than information that was not relevant (b = 4.362, z = 50.604, $p < .001$; Figure 1.2B), suggesting that Number-game sketches included much more number information than shape information (and vice-versa). In fact, recovery of irrelevant information was close to chance (.25) in shape games at 20.8% (CI: [19.4, 22.3]), and also low, although significantly higher than chance (.125), in number games at 28.2% (CI: [26.8, 29.7]). Finally, this model also revealed that recognizers were better able to recover information from sketches produced early in communication games than sketches produced towards the end of the game (b = -0.070, z = -3.819, $p < .001$), which suggests the emergence of some dyad-specific conventions over the course of the communication games.

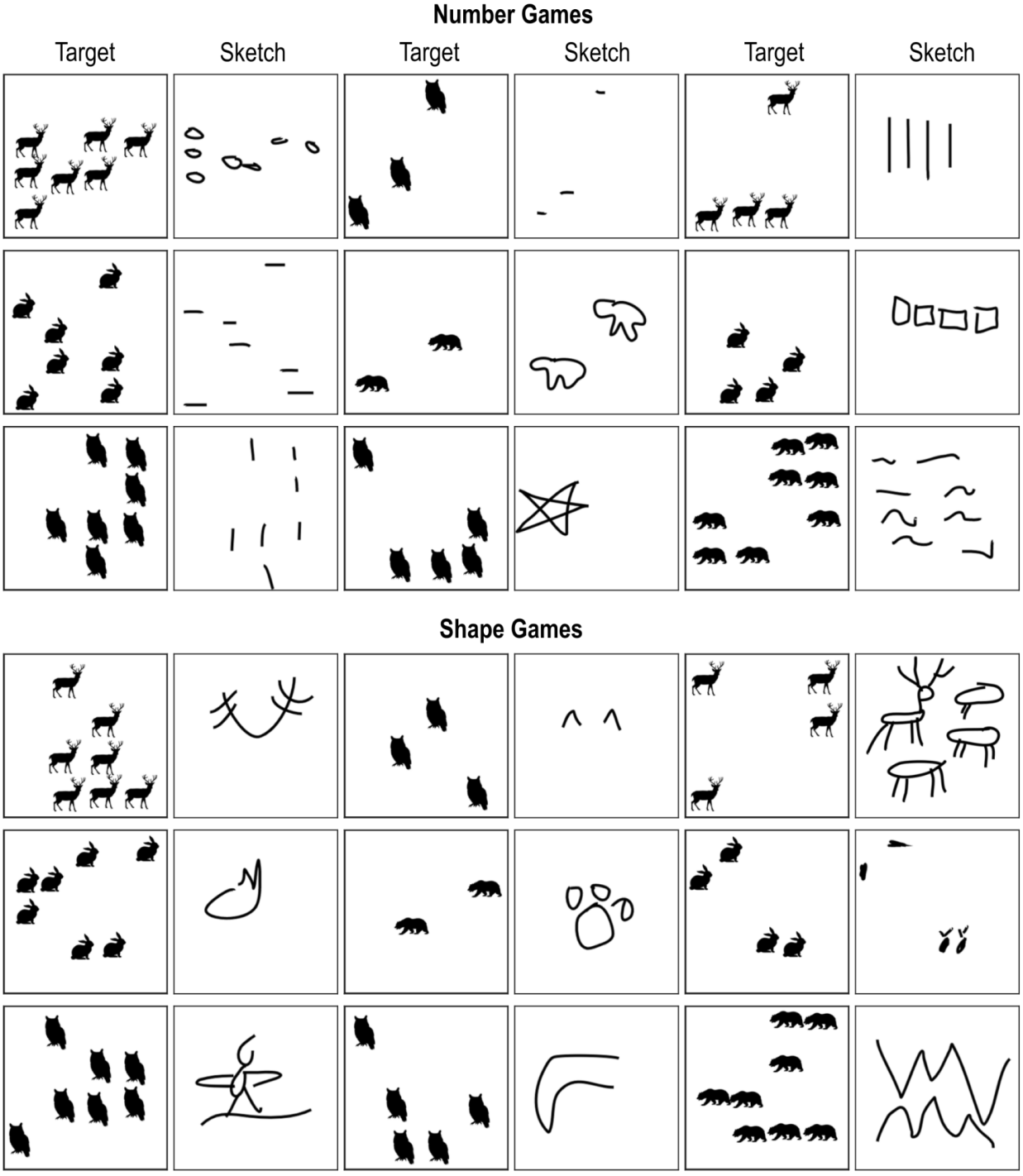


Figure 1.3. Some Example Sketches. Above. Selection of nine target arrays from number game trials, with the sketch made for each to its right. **Below**. Same as above, but shape game trials.

1.3 Discussion

In this experiment, we found that participants communicated effectively about both number and shape, when each feature was communicatively relevant. To communicate about number, they overwhelmingly relied on 1-to-1 correspondence, ignoring the shapes of individual objects (and vice versa for shape). Also, we found that larger numbers were communicated less accurately than smaller ones. This is consistent with prior work that has found that larger cardinalities are associated with noisier magnitude representations, particularly under speeded conditions when counting is impractical. Finally, some dyads generated novel strategies for encoding number and shape, as evidenced by the fact that the representations they used were not interpretable to independent "recognizers". For example, some participants appear to have taken advantage of the correlation between spatial configurations and cardinality in order to create depictive, configural representations of number.

These results are interesting for a number of reasons. First, despite the fact that all of the participants in our studies were fully numerate adults who were familiar with at least one numerical notation system (i.e., the Arabic numerals) it was clearly not trivial to create arbitrary representations of number *de novo*. This result provides experimental evidence that creating novel conventions for number is a qualitatively harder task than creating conventions for representing familiar objects. It also establishes that this method can be used to explore how participants might overcome this challenge. Second, the results are interesting because they demonstrate that the strategies preferred by participants align with the most common early strategies used by human ancestors - i.e., 1-to-1 correspondence that abstracts away from the identity of the items being enumerated, and using correlations in the environment as a proxy for cardinality. Participants created arrays of dots that ignored shape information almost universally,

similar to systems like those attested in Greek, Hittite, Cretan, Aramaic, Mayan, etc. Also, although 1-to-1 was readily available and easy to deploy, some dyads nevertheless innovated, used configural cues, and produced forms that could not be decoded by others, potentially compatible with the creation of novel conventions.

In Experiment 2, we explored this last finding, and began to probe how participants might make use of publicly shared correlates of number to transcend the reliance on 1-to-1 correspondence. The history of numerical symbols indicates that 1-to-1 representations sometimes evolve over time into more arbitrary conventions by exploiting visual regularities (Chrisomalis, 2020). For example, the way marks are consistently arranged on each face of a 6-sided die allows for rapid and efficient decoding and communication of numerical information. Such configural cues are also exploited in the design of the abacus to facilitate the recognition and manipulation of larger cardinalities (Frank & Barner, 2012; Srinivasan et al., 2018), and in some early writing systems, as described in the Introduction. In each case, rather than counting each mark, people can map a recognizable visual pattern onto a specific cardinality. Because humans are adept at exploiting incidental correlations to communicate quantity (Cooperrider & Gentner, 2019), it is possible that the presence of visual regularities of sets in the world could facilitate the creation of compressed representations that exploit them. We explored this in Experiment 2 by creating contexts that enhanced the availability of configural cues, while also increasing the incentive to innovate by presenting participants with larger sets.

Experiment 2

In Experiment 2, we manipulated the visual regularity of sets to ask whether this might lead participants to produce sketches that were more compressed (i.e., using fewer strokes to

describe the same number of objects), rather than relying on 1-to-1 correspondence. Specifically, we manipulated whether sets of animals were arranged in a regular grid pattern with columns of 6 animals each, or were instead arranged in an irregular configuration, similar to Experiment 1. We reasoned that regular configurations might promote compression, e.g., by allowing participants to replace complete columns of dots with vertical lines, or rows with horizontal lines, or to use shapes (like rectangles) to indicate which grid of dots is intended. In addition, we shifted the range of cardinalities that were tested from 1-8 (in Experiment 1) to sets that were as large as 20 (in Experiment 2). We reasoned that, given that 1-to-1 is a common strategy for representing smaller numbers, participants might only have incentive to create compressed representations when the cost of 1-to-1 grows, as with larger numbers. Moreover, because people in serial reproduction experiments implicitly introduce regularity into symbolic systems that render those systems easier for others to learn (e.g., Kirby, Cornish, & Smith, 2008; see Wang et al., 2023 who show this with dot arrays similar to our stimuli), we expected stimuli that already feature visual regularity to be easier for sketchers not only to communicate about, but also to propose compressed forms that viewers could accurately decode.

2.1 Methods

2.1.1 Participants

We recruited 200 participants on AMT, as in Experiment 1. Of the resulting 100 games, 41 were excluded per our pre-registered criteria: 35 for scoring below our accuracy threshold of 50% (same as Experiment 1)[1], and 6 for using existing numerical symbols. This left 29 irregular

[1] This somewhat high exclusion rate might be explained by some participants missing the part of the task instructions which explained that the positions of the arrays were independently

condition games and 30 regular condition games. Of the remaining participants, 113 answered the optional demographic survey (50 female; M_{age} = 36.46 years, SD_{age} = 11.07 years). Finally, an error in recording reaction times affected 3 of the remaining games. While analyses reported here reflect these exclusions, results did not differ when data from these 3 games were included.

2.1.2 Design, Materials, and Procedure

Materials. Stimuli were generated as in Experiment 1, with three differences. First, only larger cardinalities (15-20) were used, rather than smaller ones (i.e., 1-8). Second, these cardinalities were presented in two different ways: For games in the Irregular condition, the spatial layout of animal silhouettes was again generated by sampling locations for each item at random as in Experiment 1. For games in the Regular condition, however, the animal silhouettes were arranged in a rectangular grid pattern, with each column containing 6 items and the rightmost column containing the remainder (e.g., 15 had a remainder of 3 animals in the rightmost column; Figure 4A). Finally, only three of the previous animal types were used (i.e., bear, deer, and owl, but not rabbit).

Procedure. The procedure was identical to Experiment 1, with two important differences. First, participants saw each animal-cardinality combination twice, once in the first half of the experiment (18 trials) and once in the second half (18 more trials), rather than only once as in Experiment 1. Further, each block of 6 trials contained one instance of every cardinality as the target. Apart from these constraints, the presentation order of stimuli was

randomized for the sketcher and viewer. For example, some sketchers always drew 1-4 marks, corresponding with the position of the target array among the distractors.

randomly shuffled within each block of 18 trials. Second, sketchers and viewers saw each array in the same spatial configuration as each other, allowing configural information to be leveraged for communication. This was impossible in Experiment 1, as spatial configuration was randomized independently for each partner. Furthermore, for each sketcher-viewer dyad, arrays of each cardinality were consistently presented in the same spatial configuration across trials, making it possible for sketchers to use the same configural cues to communicate about particular cardinalities across trials – and thus to develop ad hoc graphical conventions. We did not deem it necessary to explicitly inform participants of this fact, based on our observation that in Experiment 1 most participants in the number condition produced drawings that preserved information about how objects were arranged in the target array, even when this information was irrelevant (because the spatial configuration of objects shown to the viewer was different).

2.2 Results[2]

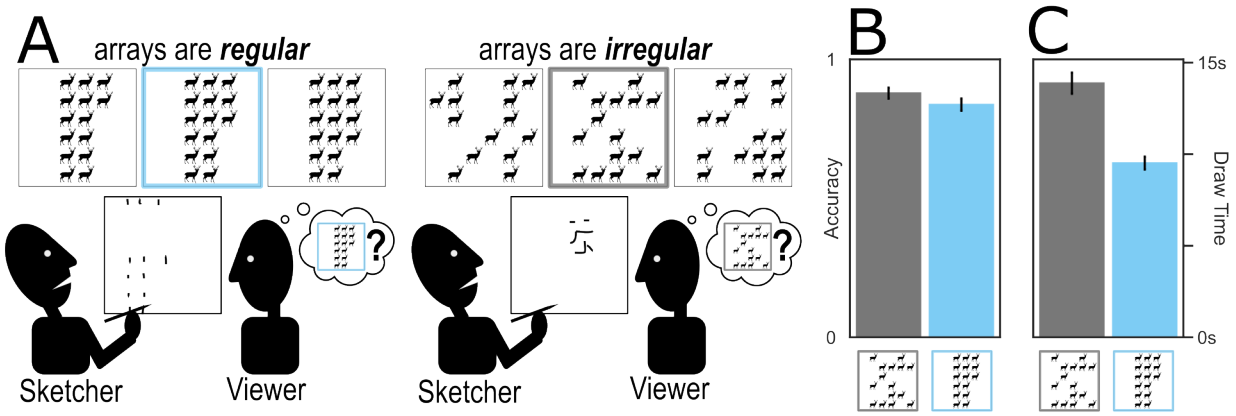


Figure 1.4. A. In the regular condition (left), all arrays were organized into columns of 6, with a remainder column on the right. In the irregular condition (right), arrays were organized randomly, as in Experiment 1. **B.** Viewers were slightly less accurate when decoding sketches of regular arrays (error bars are 95% CI). **C.** Sketchers made sketches in much less time when encoding regular arrays than irregular ones (error bars are 95% CI).

[2] Two pre-registered analyses are not reported. The first was a spatial clustering analysis to measure visual chunking, and the second was a classification task asking online participants to infer sketch strategies, as in Experiment 1; both proved to be uninterpretable.

2.2.1 Communicative Efficiency. We first investigated communicative efficiency, which we measured in three ways: (1) how accurate pairs were, (2) how long sketchers took to make their sketch, and (3) how long viewers took to submit their guess. We found that dyads in both conditions achieved high accuracy (Regular: 83.7%, CI: [81.5, 85.9]; Irregular: 87.8%, CI: [85.9, 89.8]). We constructed a binomial linear mixed-effects model to predict communicative success similar to Experiment 1, but also included an interaction term between regularity and cardinality. As in Experiment 1, this model revealed that dyads were more accurate when communicating small numbers than large numbers (b = -0.122, z = -2.020, p = 0.043), more accurate in later trial blocks (b = 0.421, z = 9.760, p < .001), and more accurate in the Irregular condition than in the Regular condition (b = -4.604, z = -3.173, p = 0.002). Sketchers took less time to communicate regular arrays than irregular arrays (b = -4.363, t = -3.23, p = 0.002), although there was no difference between conditions in the time that viewers took to make their guesses (b = 22.1036, t = 1.349, p = 0.178). This result could reflect the fact that sketchers could estimate arrays more quickly when organized into groups (Mandler & Shebo, 1982; Atkinson, Francis, & Campbell, 1976; Van Oeffelen & Vos, 1982; Starkey & McCandliss, 2014; Ciccione & Dehaene, 2020) or that sketches of regular arrays can be drawn more quickly than sketches of irregular arrays. Together with the model predicting accuracy, these results suggest that sketchers presented with regular visual arrays were quicker than those shown irregular arrays, but that they communicated number information less reliably (Figure 1.4B & 1.4C). Additionally, while viewers were slower and less accurate as target arrays grew in size, this effect was smaller in the Regular condition than in the Irregular condition (response time:b = -1.709, t = -2.295, p = 0.022; accuracy: b = 0.236, z = 2.942, p = 0.003). These results suggest that regular arrangements of objects provide a

more scalable representation of large numbers, allowing viewers to decode the meaning of visual representations faster despite the increasing size of sets and complexity of representations.

2.2.2 Communicative Strategy. We next investigated the degree to which participants used 1-to-1 or more compressed strategies to communicate about number, such as relying on configural cues. As in Experiment 1, we did so by computing stroke ratios, which should be close to 1 when participants use a 1-to-1 strategy and smaller than 1 when they use compressed strategies. Unlike the Number condition of Experiment 1, in which participants used a 1-to-1 strategy to communicate about number in 77.2% of trials, participants in Experiment 2 used much less frequently in both the Irregular condition (17.5%, CI: [15.2, 19.8]) and the Regular condition (16.8%, CI: [14.5, 19.0]). Instead, compressed strategies were common: whereas Number condition participants in Experiment 1 employed compression in only 2.9% of sketches, participants in Experiment 2 employed compression frequently in both the Irregular (77.0%, CI: [74.5, 79.6]) and Regular (78.2%, CI: [75.7, 80.6]) conditions.

We constructed a linear mixed-effect model to predict the stroke ratio of each sketch, which revealed that participants employed compression to a similar degree regardless of the visual regularity of arrays they had to communicate (b = -0.065, t = -0.993, p = 0.325). Also, representations of larger cardinalities were significantly more compressed than representations of smaller ones (b = -0.014, t = -6.760, p < .001), and compression was greater in later trial blocks (b = -0.013, t = -6.393, p < .001). Together, these results suggest that as participants gained experience with the task, they became more effective at encoding number in their sketches, and that the regularity of spatial arrangements had no effect on either the choice of strategies or the amount of compression employed when participants did depart from a 1-to-1 strategy.

2.3 Discussion

Experiment 2 asked whether visual regularities in an array of objects help participants create compressed representations of number. While participants in Experiment 1 were reluctant to communicate about number information using strategies other than 1-to-1 correspondence, sketches in Experiment 2 were much more compressed, relying substantially less on 1-to-1 correspondence. Two features of Experiment 2 may explain this finding. First, arrays in Experiment 2 were presented in configurations - either regular or irregular - that were shared between the sketcher and the viewer. Second, the sets in Experiment 2 were larger than those in Experiment 1. Compatible with the idea that this may have promoted compression, participants in Experiment 2 were significantly more likely to use compressed representations for larger sets than for smaller ones. This suggests that the incentive to create compressed representations of number may be greatest when individuals are asked to communicate about large quantities - a fact that is compatible with the historical record, in which systems often used 1-to-1 strategies for small numbers, but compressed conventions for larger ones.

In both Experiments 1 and 2, we provided participants with numerical arrays and asked them to spontaneously generate novel symbols that might be used to represent number. These studies found evidence that participants are eager to use shared correlates of number to communicate, but also that generating conventions to represent number *de novo* is no easy task in such contexts. However, as noted in the Introduction, evidence from the historical record indicates that conventions for labeling number and numerical measures rarely emerged from purely arbitrary innovations, and very frequently arose by processes of *ad hoc* comparison to ubiquitous, publicly available objects, like body parts. Also, although numeral systems

sometimes exploit configural cues to represent number, most systems do not, perhaps because configural cues are not reliable properties of the things that humans wish to communicate about.

In Experiment 3, rather than burdening participants with the need to create entirely novel representations from scratch, we probed their interest in exploiting publicly shared correlates of number by providing them with objects that could act as candidate conventions, by virtue of their numerically relevant shapes (e.g., a pair of cherries, or a four-leaf clover). This allowed us to not only explore this idea, but also to test whether, when using such objects to communicate, participants show evidence of using combinatorial rules akin to those found in numeral systems in the attested historical record.

Experiment 3

Experiment 3 asked how, if at all, communicators would exploit numerical features of familiar objects in their shared environment to express number, and whether their use of such features might facilitate the use of combinatorial rules. The use of *ad hoc* comparison to convey quantity is frequently attested in human languages (Epps, 2006; Cooperrider & Gentner, 2019), as is the combination of small number words to express larger quantities (Comrie, 2011). To facilitate the adoption of such *ad hoc* comparisons, we allowed participants to communicate using visual tokens that featured conspicuous numerical properties (e.g., a pair of cherries, a four leaf clover, etc.). These visual tokens were included on a keyboard that participants used to communicate with a partner, and permitted the combination of multiple, concrete images to express larger numbers. We asked whether participants would persist in using a 1-to-1 correspondence strategy (e.g., 1 clover or pair of cherries per item), or instead use the numerical

information available in these images as symbols for number. Also, we asked whether participants would adopt rules for combining these shapes, and if so, which rules.

3.1 Methods

3.1.1 Participants

We recruited 110 participants from Prolific. Twenty-nine games were excluded according to pre-registered criteria of finishing all trials (10 games) and having a minimum of 50% accuracy (19 games), leaving 62 participants paired in 31 games (27 female; M_{age} = 26.48 years, SD_{age} = 8.16 years). All participants provided informed consent as per the IRB.

3.1.2 Design, Materials, and Procedure

Materials. Stimuli were generated according to the same process as in Experiments 1 and 2, but included both small and large cardinalities (1-16), and contained only dots instead of different kinds of animals. This was because based on the results of Experiments 1 and 2, the exact animals appeared to no longer be theoretically relevant. Rather than a sketchpad, participants were presented with a virtual keyboard consisting of four buttons. Each button featured a different natural object, each of which had a different number of conspicuous physical features ranging from 1 to 4 **(**Figure 1.5A**).** These were an apple, a pair of cherries, three connected oak leaves, and a four-leaf clover. There was also a 'delete' key. While no explicit limit was told to participants, the task interface limited messages to no more than 30 tokens – a limit which was never reached.

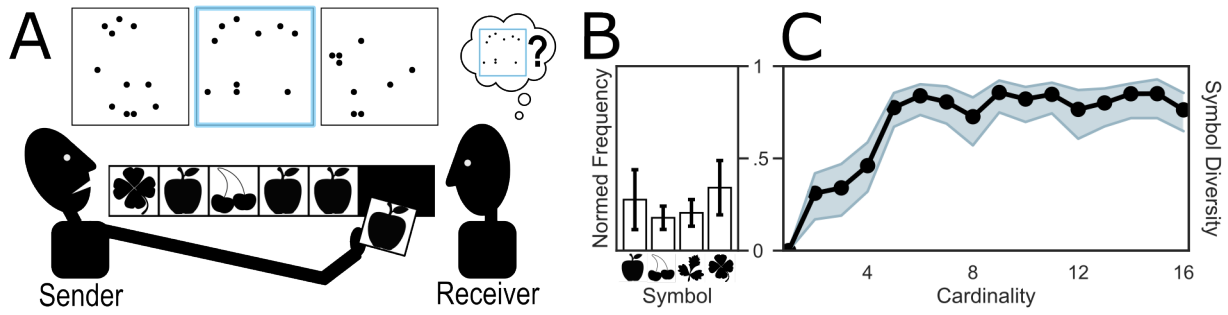


Figure 1.5. A. In Experiment 3, participants communicated about arrays of dots that varied in quantity. Rather than a sketchpad, they communicated using a keyboard of existing shapes. B. Relative frequency of each of the 4 shapes across all messages (error bars are 95% CI). C. Simpson's diversity index of messages created to represent sets of each cardinality.

Procedure. The procedure was similar to Experiments 1 and 2, in that a *sender* (analogous to the *sketcher*) sent messages to a *receiver* (analogous to the *viewer*) to communicate one image out of a larger set of images. However, it was different in two respects. First, all 16 cardinalities appeared once in the first half of the experiment, and once again in the second half, for a total of 32 trials. Second, we reduced the total number of images presented on each trial from 4 to 3 (i.e., 1 target + 2 distractors). This was to discourage participants from trying to communicate the ordinal position of the target image (e.g., 2nd from the left) by using the ordinal position of the four response keys, which were also arranged horizontally. Such a strategy would result in chance performance, as the ordinal position of images was randomized independently between partners. Like Experiment 1, the spatial configuration of dots within each array was also independently randomized for each partner, to preclude reliance on configural cues, as well as between every two occurrences of that array within the task.

3.2 Results

3.2.1 Communicative Efficiency. We measured communicative efficiency in three ways: (1) how accurate pairs were, (2) how long senders took to make their message, and (3) how long receivers took to submit their guess. To do this, we constructed linear mixed effects regressions, modeling the effects of cardinality and trial number on each measured variable. Overall accuracy was slightly lower than in Experiments 1 (91.9%) and 2 (85.7%), at 78.7%. As in Experiments 1 and 2, participants improved at the task as they gained experience with it: dyads became more accurate over successive trials ($b = 0.062$, $z = 6.154$, $p < .001$), senders took less time to compose their messages ($b = -0.180$, $t = -3.119$, $p = 0.002$), and receivers took less time to make their guesses ($b = -0.207$, $t = -2.374$, $p = .018$). Also similar to Experiments 1 and 2, larger cardinalities were communicated less accurately than smaller ones ($b = -0.147$, $z = -7.13$, $p < .001$), and more slowly both by senders ($b = 1.279$, $t = 11.3$, $p < .001$) and receivers ($b = 1.000$, $t = 5.849$, $p < .001$).

3.2.2 Communicative Strategy. We next investigated the degree to which participants used 1-to-1 or compressed strategies to communicate number. Analogous to the stroke ratio measure employed in Experiments 1 and 2, we measured *token ratio*: the ratio between the number of tokens used within a message to the cardinality of its target array. A ratio close to 1 would suggest the use of 1-to-1 correspondence, while a ratio less than 1 would suggest the use of a more compressed strategy. By this measure, participants in Experiment 3 used 1-to-1 strategies in 14.9% of trials, similar to the 17.1% of trials measured in Experiment 2, but much less than the 77.2% of trials in the Number condition of Experiment 1. Removing trials in which the target cardinality was 1 (1/16 of all trials), where the use of 1-to-1 would thus be indistinguishable from other strategies, this proportion was only 9.2% - significantly lower than in Experiment 2

(Fisher's exact test; $p < .001$). Instead, compressed strategies were evident in 84.6% of trials, suggesting that the use of 1-to-1 correspondence was relatively infrequent overall and limited to a small number of games. Finally, senders were also significantly more likely to compress information when they were communicating larger cardinalities than smaller ones ($b = -0.032$, $t = -7.286$, $p < .001$), and this did not change over successive trials ($b = 0.001$, $t = 0.540$, $p = 0.589$). This distribution of strategies suggests that the presence of objects that support *ad hoc* comparison helps communicators to employ strategies that bypass direct 1-to-1 correspondence between tokens and objects.

An exploratory linear model predicting accuracy from token ratio revealed that compressed forms did not significantly contribute to higher accuracy ($b = -0.428$, $z = -1.175$, $p = 0.240$), and may have reduced accuracy as numbers became larger ($b = 0.195$, $z = 3.166$, $p = 0.0015$). This may be because larger numbers could be expressed in a greater number of ways, as they appear to have been in senders' messages. For example, while there was only one attested form across all games for expressing the number *1* (that is, a single *apple* token), there were 5 unique forms for expressing the number *2*, and 19 forms for expressing the number *12*. Furthermore, the share of trials exhibiting the most commonly used form for each number dropped as numbers increased. While all 62 occurrences of the number *1* were represented by a single apple, the most common expression for the number *2* was used 51 times, and the most common expression for *12* was used only 29 times. This is reflected in the Simpson's diversity index of messages used to represent sets of each cardinality, which is close to 0 when the same expression is always used to represent a cardinality, and close to 1 when many different expressions are used to represent a cardinality, with equal frequency. In our data, this index is closest to 0 for the smallest cardinalities, and is close to 1 for large cardinalities (Figure 1.5C).

3.2.3 Rule Use. We next performed several exploratory analyses to determine which strategies were employed. We first asked whether participants made use of different kinds of rules in their messages. To do this, we measured the Levenshtein distance[3] (Levenshtein, 1966) between messages that senders produced and each of four model strategies. For any given game, the distance of every message from a model system was summed over all trials, providing a measure of how closely the sender in that game adhered to the use of one strategy or another. Finally, the distance of each game to one or another strategy was compared, and each game was assigned the strategy of the model to which it had the shortest distance.

The four possible strategies we assessed were: 1-to-1, cumulative-additive, place-value, and a 'single token' strategy. The 1-to-1 strategy entails repeating one shape as many times as there were objects in the target set (e.g., using 4 apples to represent a set of 4). The cumulative-additive strategy involved summing tokens to communicate a number, but drawing on the more complex shapes to represent the numbers 2, 3, and 4 (e.g., using 2 cherry pairs plus 1 apple to represent 5). Under this strategy, expressing numbers greater than the largest image (the four leaf clover) involved repeating that image as many times as 4 could be divided into the number, with a remainder expressed via other shapes with smaller cardinal meanings. For example, the number 10 might be expressed as *clover-clover-cherries*, or *442*. The third strategy, which we expected to be quite unlikely, was a place-value system, similar to Arabic notation. In a system with only 4 possible shapes to use as symbols, the second place value represents multiples of 4, such that a shape representing 1 (e.g., apple) produces a value of 4, while a shape representing 2 (e.g.,

[3] A measure of string distance defined as the number of character insertions and deletions required to turn one string into another, often used in studying artificial symbol systems (e.g. Cornish et al., 2017; Vokey & Brooks, 1992).

cherry) produces a value of 8, such that a string like *cherry-apple* represents 9. Finally, we modeled an ordinal system, where the characteristic cardinality of the shape chosen represented the ordinal position of the target among the distractors, ignoring their cardinality entirely (i.e., 1 apple for 1st set from the left, a pair of cherries for 2nd, etc.).

This analysis suggested that most games relied on a cumulative-additive system (17 games), while a smaller number relied primarily on 1-to-1 correspondence (3 games). A cumulative additive system is also reflected in the frequency of each token (apple, cherries, oak leaves, and clover): the apple and clover were the most common tokens (Figure 1.5B). Preliminary classifications suggested that 6 additional games were closest to the ordinal strategy, and 5 to the place-value system. However, on closer, qualitative inspection, the strategies in these latter two groups of games may reflect a naïve grouping strategy, where groups of objects in the target image are represented in order as they're viewed, so that 8 objects arranged into groups of 1, 2, 1, 3, and 1 are represented as 12131. Given the observed distribution of strategies, the cumulative-additive strategy may be the only strategy reliably used by participants across our studies to feature a regular syntax – perhaps enabled by the change of communicative medium.

3.2.4 Order. Finally, one property of the cumulative-additive system that was common across games was the arrangement of tokens in a decreasing order, where 'smaller' tokens were always to the right of larger tokens (e.g., *442* is decreasingly ordered but *424* or *244* are not). Overall, 87.7% of messages were decreasingly ordered in this way, a number that did not significantly differ between correct and incorrect trials (88.1% and 86.3%). To understand whether this trend should be expected by chance, we generated a random message for each trial, such that the sum of all tokens in the participant's message and the random message were the same. We then used

a permutation test to ask whether our random messages exhibited decreasing order as often as participants' messages, and found that participants' messages were significantly more likely to be decreasingly ordered ($t = 8.24$, $p < .001$). This preference was found across most games, and may have facilitated the process of reading those messages. Exploratory t-tests indicate that the reaction time of receivers was much faster when confronted with decreasingly ordered messages from their sender ($t = 5.21$, $p < .001$). More surprisingly, a similar trend holds for the time it took senders to construct their messages, as decreasingly ordered messages were also faster to make ($t = 5.742$, $p < .001$), though this trend may owe itself to faster reactions and better performance of those participants who also opted to use this strategy, rather than an effect of the strategy per se.

3.3 Discussion

In Experiment 3, we provided participants with pictures of familiar objects that had stable numerical properties (e.g., a pair of cherries, a four leaf clover), and asked whether they would exploit these properties to communicate about number. We found that few participants persisted in using 1-to-1, and that most instead used the features of provided objects to compress their representations of number. The prevalence of compressed strategies (84.6%) was slightly greater than in Experiment 2 (77.6%). Unlike Experiment 2, however, the modality of Experiment 3 required participants to arrange a set of discrete shapes in a linear order. Given this constraint, many participants invented a numerical base out of the 4-leaf clover shape with the result that most messages in shared two features that are characteristic of historically attested numeral systems. One feature was the use of a cumulative-additive structure, taking advantage of the highest available base (e.g., expressing 5 as *clover* + *apple*, rather than as *cherries* + *cherries* + *apple*). The other characteristic feature was the ordering of tokens from largest to smallest,

similar not only to the familiar Arabic numeral system, and also to the majority of linear numeral systems in historical record (Chrisomalis, 2020).

General Discussion

Across three experiments, we investigated how humans create and combine symbols to express number. In Experiment 1, we found that when participants were asked to communicate about number to a partner, they often created sketches that used a 1-to-1 correspondence strategy: For each object in a set, they generally created one corresponding dot, mark, or sketch of that thing. However, they sometimes used configural cues to express number, and in some cases, dyads used representations that could not be decoded by independent participants, potentially compatible with the creation of new conventions. In Experiment 2, we directly explored this finding, and found that when configural cues were reliably available across trials participants used 1-to-1 correspondence to communicate number much less frequently, and often used compressed representations that exploited configural cues. In Experiment 3, rather than requiring participants to invent new conventions *de novo*, we asked whether they would exploit publicly shared correlates of number by providing them with shapes that could act as candidate conventions (e.g., a pair of cherries, or a four-leaf clover). Here we found that participants rarely used 1-to-1 correspondence in favor of compressed representations, and also that when they combined shapes to communicate number, they often used a descending cumulative-additive structure, using the largest available "symbol" as a base (i.e., the clover).

These results suggest two main conclusions. First, the problem of creating entirely novel representations of number may be uniquely difficult when proxies for number, like shape, are not available in the communicative context. Whereas Shape game participants in Experiment 1

readily created conventions that used minimal strokes and were hard for naïve "recognizers" to decode (as in Garrod et al., 2007; Hawkins et al., 2023), participants in Number games invented conventions much less often, as reflected both by their persistent use of one-to-one correspondence and the relative ease with which recognizers identified the target numbers communicated by sketches. This was despite the fact that our participants were numerate adults who were familiar with western Arabic numerals. However, a second conclusion is that participants readily create compressed representations of number when non-numerical proxy representations are available in the context. Although most participants in Experiment 1 relied on 1-to-1 correspondence, some tried to communicate number by preserving the configural cues contained within stimuli. In Experiment 2, when these configural cues were more robust and reliable (i.e., arranging objects into rows in a consistent manner), participants frequently departed from the use of 1-to-1 strategies. Finally, this reliance on shape information was strongest when shapes of familiar objects - like apples, or clovers - could be combined to express number (Experiment 3).

As noted in the Introduction, humans often use ubiquitous and publicly available objects like hands, feet, and other body parts to count, and often use labels for these parts to subsequently name different cardinalities (Dahl, 1981; Epps, 2006; Heine, 2004; Rischel, 1997). Also, in at least some languages, small sets are labeled using names for common objects that feature specific numbers of things - like the number of seeds in a fruit, the number of eyes on one's head, or the number of fingers and toes on one's body (Epps, 2006). Similar practices of so-called *ad hoc* comparison can also be found in the history of measurement systems (Cooperrider & Gentner, 2019), but also in other cases such as color, where hues are labeled via metonymy, using the names of things in the local environment that have the same color (e.g.,

orange, lilac, gold, etc.; Casson, 1994). The logic underlying such instances of metonymy may help explain why participants in Experiments 1 and 2 did not create novel conventions as readily as in other studies of graphical communication. For example, in a study by Fay et al. (2010) some participants conveyed the concept *parliament* by initially sketching members of parliament sitting at tables with the Australian flag, but later included only the flag, which sufficed as an index of the concept even without the depiction of the people. This was possible because the concept parliament is associated with multiple imageable components which, after being used collectively to identify the intended referent, can be subsequently reduced to only one element (e.g., a flag) to create a convention. However, such a strategy is often not available in the case of number, since in a set of, e.g., four ducks, no individual duck has any feature that can alone communicate the cardinality four. The use of 1-to-1 correspondence appears to be overcome only when there exists an imageable correlate of number, such as a spatial configuration, or a shape or name that is strongly associated with a particular cardinality (e.g., eyes, clovers, etc.).

In addition to finding that participants often use physical correlates of number to communicate about large quantities, the results from this study also suggest ways in which the medium of numerical symbols may impact the form that they take. Whereas the reference sets in all experiments – as well as the communicative medium in Experiments 2 and 3 – were two-dimensional visual arrays, only in Experiment 3 did the communicative medium require a strict linear ordering of component shapes. This appears to have led participants to create combinatorial representations of number, perhaps by prompting them to reflect on their strategy for encoding the objects in the target set. This is potentially important, because previous theories of the history of number posit that abstraction occurs when representations transition from purely referential meanings towards meanings that are defined by relations between elements within the

symbolic system. New abstractions, like the rules that govern written numerals, are made possible by the affordances of the symbolic medium in which the system is instantiated - in this case a linearly ordered sequence (Overmann, 2018).

While this collection of exploratory studies provides a first step toward understanding the processes that underpin the creation of numeral systems, they had several notable limitations. One is that some of our pre-registered analyses proved intractable in the face of the data we collected. For example, a planned analysis of spatial clustering of strokes within sketches was unable to meaningfully recover coherent clusters that were apparent to the human eye (e.g., dotted lines drawn by participants), rendering results of that analysis questionable. Also, we initially planned to conduct a recognition task in Experiment 2 that paralleled Experiment 1, but found that participants gave highly divergent and often uninterpretable responses to sketches, making it unclear whether the labeling techniques we offered them were understood. Another limitation is that our methods did not make it possible to easily identify the cognitive steps involved in the creation of number representations. For example, it is possible that some participants in Experiments 1 and 2 did not even consider creating summary representations of cardinality, and instead tried to communicate the visual patterns present in arrays, rather than number.

Another limitation, common to communicative games, is that participants in our study could only communicate via the medium we provided to them, and were not allowed to directly tell partners what symbols meant. This constraint was important to precluding uninteresting strategies in our numerate participants (like creating a direct translation of existing numerals). If laboratory participants were allowed to make explicit agreements, they could easily bypass the most important communicative obstacle to creating numerals from scratch by using known

symbols as anchors for creating new ones - e.g., by explicitly agreeing to replace "5" with "%" (or some novel form). Also, we felt that our method might simulate the challenge of expressing exact number when existing symbols are absent, since this requires expressing number via some other medium or strategy. Still, most number systems almost certainly evolved through processes of *conspiracy* among community members, in which individuals worked together to create, agree upon, and teach new symbols. Exploring how such collaborative processes might work when existing numeral conventions are not available should be explored further in future work. Indeed, another limitation of our study is that all of our participants were numerate adults. Although their numeracy proved to be of little help in creating novel systems, it remains an interesting question whether innumerate adults or young children might deploy different strategies - a question we are currently exploring.

In summary, results from three experiments suggest that when numerate adults are tasked with devising novel ways to communicate about number, they often default to using 1-to-1 correspondence, but also readily exploit visual correlates of number to efficiently convey representations (e.g., configural cues or objects that canonically appear in certain numbers). Features of the medium of communication may also influence the creation of conventions by limiting degrees of freedom, and by prompting communicators to spontaneously organize their messages in novel ways.

Acknowledgement

Chapter 1, in full, is a reprint of the material as it appears in Holt, S., Fan, J. E., & Barner, D. (2024). Creating ad hoc graphical representations of number. *Cognition, 242*, 105665. The dissertation author was the primary researcher and author of this paper.

Declarations: Competing Interests, AI, Funding

The authors have no interests that may influence or appear to have influenced the above work. AI was not used in the writing process. Funding sources were: NSF Award #2000827 to DB; NSF CAREER #2047191 to JEF.

Data Availability

All pre-registrations, materials, data, and analyses are available at the authors' Github project repository: https://github.com/cogtoolslab/improvised_numerals

References

Atkinson, J., Campbell, F. W., & Francis, M. R. (1976). The magic number 4±0: A new look at visual numerosity judgements. *Perception, 5*(3), 327-334.

Barner, D. (2017). Language, procedures, and the non-perceptual origin of number word meanings. *Journal of Child Language, 44*(3), 553-590.

Beller, S., Bender, A., Chrisomalis, S., Jordan, F. M., Overmann, K. A., Saxe, G. B., & Schlimm, D. (2018). The cultural challenge in mathematical cognition. *Journal of Numerical Cognition, 4*(2).

Bender, A., & Beller, S. (2011). Fingers as a tool for counting–naturally fixed or culturally flexible? *Frontiers in Psychology, 2*, 256.

Bender, A., & Beller, S. (2012). Nature and culture of finger counting: Diversity and representational effects of an embodied cognitive tool. *Cognition, 124*(2), 156-182.

Brannon, E. M., & Merritt, D. J. (2011). Evolutionary foundations of the approximate number system. *Space, time and number in the brain*, 207-224.

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

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

Casson, R. W. (1994). Russett, rose, and raspberry: the development of English secondary color terms. *Journal of Linguistic Anthropology, 4*(1), 5-22.

Changizi, M. A., & Shimojo, S. (2005). Character complexity and redundancy in writing systems over human history. *Proceedings of the Royal Society B: Biological Sciences, 272*(1560), 267-275.

Chrisomalis, S. (2004). A cognitive typology for numerical notation. *Cambridge Archaeological Journal, 14*(1), 37-52.

Chrisomalis, S. (2010). *Numerical Notation: a Comparative History*. Cambridge University Press.

Chrisomalis, S. (2020). *Reckonings: Numerals, Cognition, and History*. MIT Press.

Christiansen, M. H., & Kirby, S. (2003). Language evolution: Consensus and controversies. *Trends in Cognitive Sciences, 7*(7), 300-307.

Ciccione, L., & Dehaene, S. (2020). Grouping mechanisms in numerosity perception. *Open Mind, 4*, 102-118.

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

Conant, L. L. (1905). The beginnings of counting. *School Science and Mathematics, 5*(6), 385–394. https://doi.org/10.1111/j.1949-8594.1905.tb04509.x

Cooperrider, K., & Gentner, D. (2019). The career of measurement. *Cognition, 191*, 103942.

Cornish, H., Dale, R., Kirby, S., & Christiansen, M. H. (2017). Sequence memory constraints give rise to language-like structure through iterated learning. *PloS One, 12*(1), e0168532.

Dahl, O. C. (1981) Austronesian numerals. In R. A. Blust (Ed.), *Historical Linguistics in Indonesia, part I* (pp. 46–58). Jakarta: Universitas Atma Jaya.

Damerow, P. (1996). Number as a second-order concept. *Science in context, 9*(2), 139-149.

Dehaene, S. (1997). *The Number Sense: How the Mind Creates Mathematics*. Oxford University Press.

Epps, P. (2006). Growing a numeral system: The historical development of numerals in an Amazonian language family. *Diachronica, 23*(2), 259-288.

Everett, C., & Madora, K. (2012). Quantity recognition among speakers of an anumeric language. *Cognitive Science, 36*(1), 130-141.

Fay, N., Garrod, S., Roberts, L., & Swoboda, N. (2010). The interactive evolution of human communication systems. *Cognitive Science, 34*(3), 351-386.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences, 8*(7), 307-314.

Frank, M., Everett, D., Fedorenko, E., & Gibson, E. (2008). Number as a cognitive technology: Evidence from Pirahã language and cognition. *Cognition, 108*(3), 819–824.

Frank, M. C., & Barner, D. (2012). Representing exact number visually using mental abacus. *Journal of Experimental Psychology: General, 141*(1), 134.

Gallistel, C. R., Gelman, R., & Cordes, S. (2006). 12 The Cultural and Evolutionary History of the Real Numbers. *Evolution and Culture*, 247.

Garrod, S., Fay, N., Lee, J., Oberlander, J., & MacLeod, T. (2007). Foundations of representation: where might graphical symbol systems come from?. *Cognitive Science, 31*(6), 961-987.

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.

Gordon, P. (2004). Numerical cognition without words: Evidence from Amazonia. *Science, 306*(5695), 496–499.

Hammarström, H. (2010). Rarities in numeral systems. In J. Wohlgemuth & M. Cysouw (Eds.), Rethinking universals: How rarities affect linguistic theory. *Berlin: Mouton De Gruyter*, 11–60.

Hatano, G., & Osawa, K. (1983). Digit memory of grand experts in abacus-derived mental calculation. *Cognition, 15*(1-3), 95-110.

Hawkins, R. D., Sano, M., Goodman, N. D., & Fan, J. E. (2023). Visual resemblance and interaction history jointly constrain pictorial meaning. *Nature Communications, 14*(1), 2199.

Heine, B. (2004) On genetic motivation in grammar. In G. Radden & K. U. Panther (Eds.), *Studies in linguistic motivation* (pp. 103–120). Berlin/New York: Mouton de Gruyter, 103-120.

Hofer, M., & Levy, R. P. (2019). Iconicity and structure in the emergence of combinatoriality. In *Proceedings of the 41st Annual Meeting of the Cognitive Science Society*.

Høyrup, J. (1994). *In Measure, Number, and Weight: Studies in Mathematics and Culture*. SUNY Press.

Hurford, J. R. (1987). Language and Number: The emergence of a cognitive system. *Oxford: Basil Blackwell*.

Ifrah, G. (2000). *The Universal History of Numbers*. London: Harvill.

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences, 105*(31), 10681-10686.

Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. In *Soviet Physics-Doklady 10*(8), 707-710.

Malafouris, L. (2010). Grasping the concept of number: how did the sapient mind move beyond approximation. *The archaeology of measurement: Comprehending heaven, earth and time in ancient societies*, 35-42.

Mandler, G., & Shebo, B. J. (1982). Subitizing: an analysis of its component processes. *Journal of Experimental Psychology: General, 111*(1), 1.

O'Shaughnessy, D. M., Gibson, E., & Piantadosi, S. T. (2021). The cultural origins of symbolic number. *Psychological Review, 129*(6), 1442–1456.

Overmann, K. A. (2018). Constructing a concept of number. *Journal of Numerical Cognition, 4*(2), 464-493.

Pica, P., Lemer, C., Izard, V., & Dehaene, S. (2004). Exact and approximate arithmetic in an Amazonian indigene group. *Science, 306* (5695), 499-503.

Rischel, J. (1997) Typology and reconstruction of numeral systems: The case of Austroasiatic. In J. Fisiak (Ed.), *Linguistic Reconstruction and Typology* (pp. 273–312). Berlin/New York: Mouton de Gruyter.

Robson, E. (2008). *Mathematics in Ancient Iraq: A Social History*. Princeton University Press.

Saxe, G. B. (1981). Body parts as numerals: A developmental analysis of numeration among the Oksapmin in Papua New Guinea. *Child Development 52*(1), 306-316.

Saxe, G. B., & Esmonde, I. (2012). *Cultural development of mathematical ideas: Papua New Guinea studies*. Cambridge University Press.

Schmandt-Besserat, D. (1978). The earliest precursor of writing. *Scientific American, 238*(6), 50-59.

Schmandt-Besserat, D. (2010). *How Writing Came About*. University of Texas Press.

Schneider, R. M., Brockbank, E., Feiman, R., & Barner, D. (2022). Counting and the ontogenetic origins of exact equality. *Cognition, 218*, 104952.

Srinivasan, M., Wagner, K., Frank, M. C., & Barner, D. (2018). The role of design and training in artifact expertise: The case of the abacus and visual attention. *Cognitive Science, 42*, 757-782

Starkey, G. S., & McCandliss, B. D. (2014). The emergence of "groupitizing" in children's numerical cognition. *Journal of Experimental Child Psychology, 126*, 120-137.

Stigler, J. W. (1984). "Mental abacus": The effect of abacus training on Chinese children's mental calculation. *Cognitive Psychology, 16*(2), 145-176.

Van Oeffelen, M. P., & Vos, P. G. (1982). A probabilistic model for the discrimination of visual number. *Perception & Psychophysics, 32*, 163-170.

Verhoef, T., Kirby, S., & De Boer, B. (2016). Iconicity and the emergence of combinatorial structure in language. *Cognitive Science, 40*(8), 1969-1994.

Vokey, J. R., & Brooks, L. R. (1992). Salience of item knowledge in learning artificial grammars. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(2), 328.

Wang, Y., Lew, T. F., Brady, T. F., & Vul, E. (2023). Structured visuospatial representations revealed through serial reproduction. *Journal of Experimental Psychology: Human Perception and Performance*.

Whalen, J., Gallistel, C. R., & Gelman, R. (1999). Nonverbal counting in humans: The psychophysics of number representation. *Psychological Science, 10*(2), 130-137.

Williams, F. E. (1940a). Natives of Lake Kutubu, Papua. *Oceania, 11*(2), 121–157. https://doi.org/10.1002/j.1834-4461.1940.tb00281.x

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

Supplemental Materials

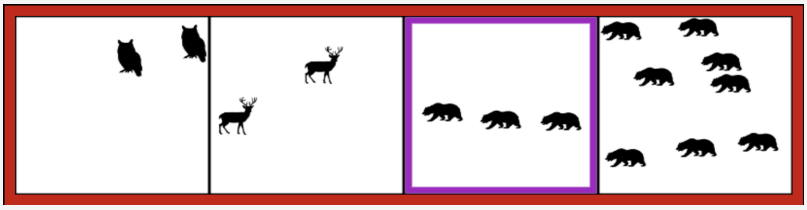
Three additional analyses were pre-registered. Each is summarized here as S1.2, S2.1, and S2.2.

S1.1. Experiment 1, Participant Instructions.

Both sketchers and viewers were presented with the same set of instructions:

Animal Game!

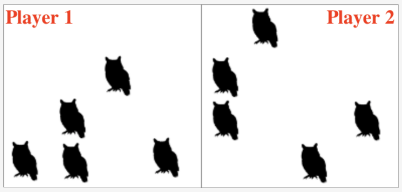
In this experiment you will play a fun communication game with another player on MTurk! Each game consists of 32 trials. On each trial both of you will see four groups of animals that vary in number and/or the kind of animal, like this:



One of you will act as the "Sketcher" and the other will be the "Viewer". Every trial will include the following steps:

STEP ONE: Both players are shown the same set of images, but in a different order. A “target” image will be highlighted to the Sketcher.

STEP TWO: The goal of the Sketcher will be to help the Viewer identify the target as quickly as possible, using the sketchpad. The Sketcher can use the sketchpad however they like, but must **NOT** draw words, arrows, numbers, or other existing symbols (although inventing new symbols is ok). Also note that the sketcher and viewer will not be shown the exact same images. For instance, a group of five owls might appear to each player in a different configuration, such as:



The Sketcher will have 30 seconds to create their message. Once they are done, they will click “submit”.

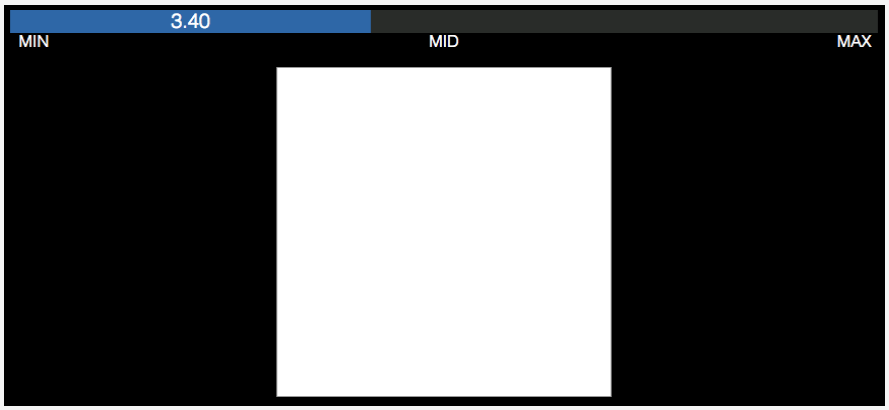
STEP THREE: The message will be shown to the Viewer, who will then click on whichever image they think it matches and click “confirm” to submit their response. **Once the "confirm" button is clicked, the response is submitted and the Viewer cannot change their selection.**

STEP FOUR: Both players will be given feedback; the Sketcher will see what the Viewer clicked (highlighted in black), and the Viewer will see the true target (highlighted in green). Then you will both automatically advance to the next trial.

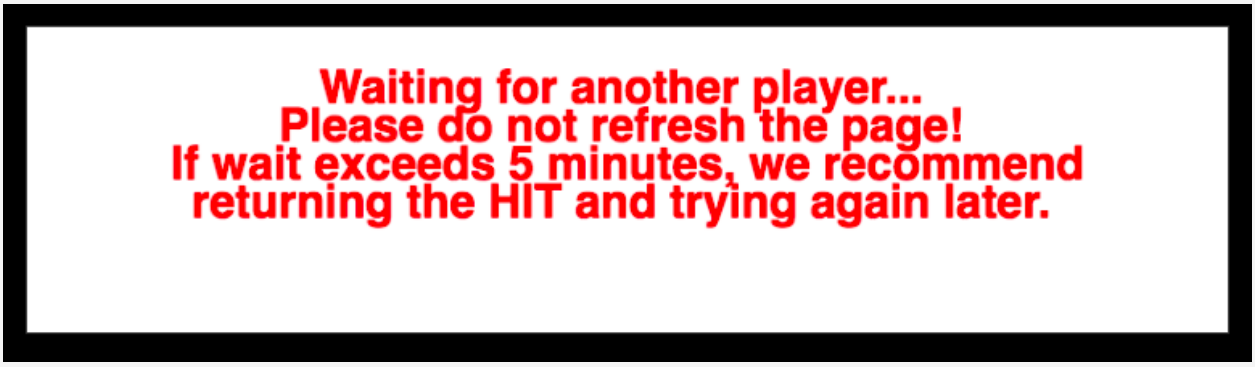
For each correct guess the Viewer makes, both of you will receive **an accuracy bonus of $0.03**. In addition, you will both receive **a speed bonus (up to $0.01) based on how fast the Viewer selects**

the correct image. In other words, **the faster the Viewer can select the correct image, the larger the bonus both of you will receive**.

However, **incorrect guesses do not receive any bonus** regardless of how quickly they were made.
The bonus meter below displays the possible total bonus (accuracy bonus + time bonus) and shrinks as the 30-second timer runs down.



A few final notes: First, since you are playing with another player, you may see this screen before the game begins:



Just hold tight. Another player should join the game within the next few minutes. If you're in the waiting room for more than 5 minutes without another player joining, please return the HIT and try again later! Otherwise, once a Viewer joins, you will be notified and need to click a button to start the game.
Second, to make sure your work is properly recognized, we ask that you be respectful of your partner and the rules of the game. If you are the Sketcher, **please do not "draw" words, arrows, numbers, or any other existing symbols**. If you are the Viewer, please do your best to be as responsive as possible so you can both earn the highest possible bonus.

Third, please DO NOT refresh the page at any point once you start the game (including the waiting page). If you refresh, you will not be able to rejoin the game you were in and you may not be able to submit the HIT.
Thank you!

In sum:

As a Sketcher, your priority is to send useful information via the sketchpad that will help the Viewer always select the correct image. **Please do not "draw" numbers, words, arrows, or any other existing symbols.**

As a Viewer, your job is to be really sure when you are making your decision so you can always select the correct image.

Please fill out this short quiz below before getting started. Good luck and have fun!

This was followed by a brief True / False quiz to check comprehension of the instructions, the questions of which were:

1. The Sketcher can click on an image to end the trial.
2. The Viewer wants to click on the image that the Sketcher has drawn.
3. The target is the image which has the purple box around it (only seen by the Sketcher).
4. Only the Sketcher can use the sketchpad to send information.
5. There are a total of 200 trials.
6. The locations of the images are the same for the Sketcher and the Viewer.
7. If an incorrect guess is made under 30 seconds, both players will still receive a "speed bonus" for completing that trial quickly.

S1.2. Experiment 1, Neural Network.

To supplement the recognition task in Experiment 1, we asked whether a neural network model of visual processing, VGG-19 (Simonyan & Zisserman, 2014), could also discern number or shape information from participants' sketches. We reasoned that the high-level feature-vector representation of each sketch from VGG-19 would encode perceptual information, but would be insensitive to any non-perceptual conventions that humans might use (e.g., drawing a carrot to represent a rabbit). In this analysis, we trained two logistic-regression classifiers to predict number or shape information, respectively, from the feature vector representation of each sketch. Both classifiers were trained and evaluated in a 5-fold cross-validated manner using the same splits as each other, and yielding a pattern of correct classifications and confusions.

This model performed similarly to naïve human recognizers in discerning shape information from sketches, but had much poorer performance in decoding number information. This suggests that human recognizers were relying not solely on low-level features (e.g., using the AMS) to recover numerical information from sketches, but were employing more abstract knowledge about quantity (e.g., 1-to-1 matching). Figure S2 shows this performance, and can be compared to Figure 1.2B in the main text.

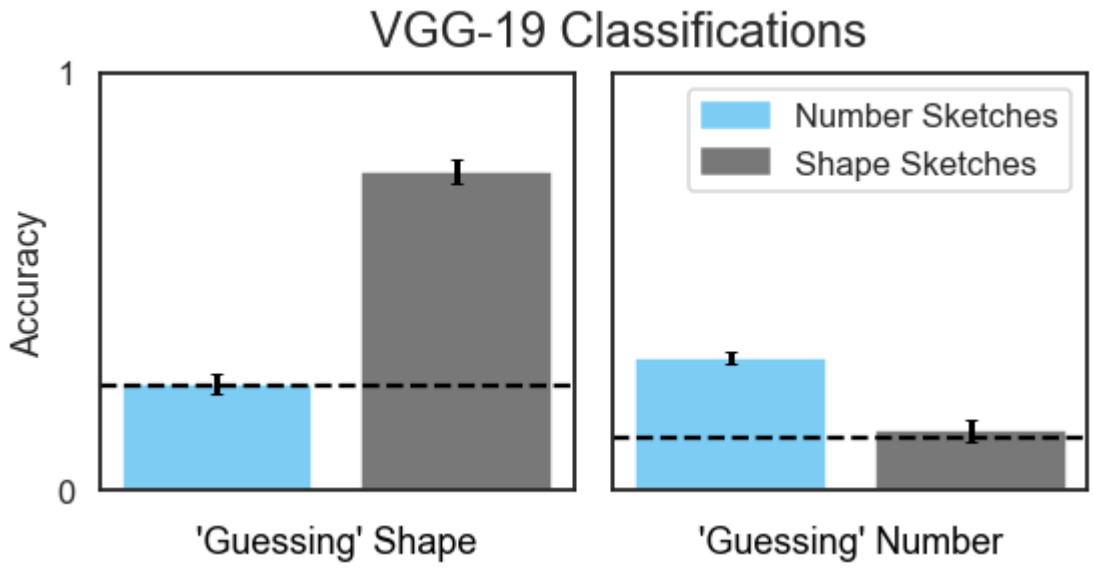


Figure S2. Classifier performance in recovering shape information (left) or number information (right) from sketches produced in Shape games (gray) or Number games (blue), using VGG-19's high-level representation of sketches. Dotted lines represent chance; error bars are 95% CI.

S2.1. Experiment 2, Classification Experiment.

In Experiment 2, we had planned to measure how well participants could classify drawings, similar to the recognition experiment from Experiment 1. In this classification experiment, participants would be asked to match the kinds of strategies used in the communication task to strategies found in historical numeral systems. We had planned to first instruct participants to label sketches from the communication game with at least one of 5 different strategies that each sketch may have employed. Each of the 5 different strategies was illustrated with one example sketch. Subsequently, during the task itself, participants classified sketches by clicking buttons that corresponded to each of the 5 strategies, and, critically, on the face of each button was the exact example sketch that had been used to illustrate the strategy. Between trials in which participants classified sketches produced during the communication task, there were also 5 'catch' trials, in which participants classified the strategies used in each of the 5 example sketches on the face of the buttons. We resolved to exclude participants if their response to each of these 5 catch trials did not include the button that displayed the exact same image. An example of the instructions and task interface can be found in the repository.

After a pilot experiment (N = 9), we decided not to continue with data collection because only 4 out of 9 participants succeeded on all catch trials, suggesting that the rate of attrition would be much higher than in the recognition task in Experiment 1. Moreover, the responses provided by the 4 participants who passed the catch trials were difficult to interpret.

S2.2. Experiment 2, Spatial Clustering.

We additionally planned on conducting an exploratory analysis of how markings were grouped within individual drawings collected in Experiment 2. The goal of this analysis was to generalize the analysis of stroke counts in Experiment 1 to accommodate clusters of strokes intended to represent the same entity in the target array. Our strategy was to employ k-means clustering to identify these groups of strokes. However, because we were unable to obtain a reliable measure of classification strategy based on human annotations (see S2.1 above), we were unable to

validate the results of this k-means clustering analysis. Furthermore, this particular approach to identifying clusters did not appear to be sensitive to groups of strokes that were related to each other based on features other than spatial proximity, such as forming a line (Figure S3).

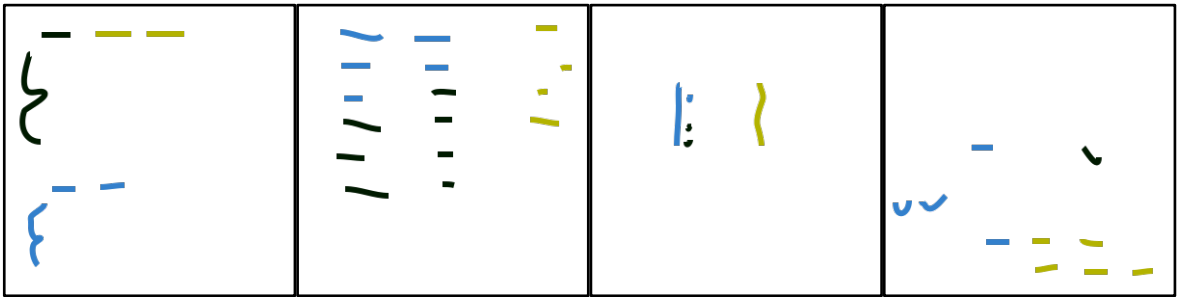


Figure S3. Four example sketches from Experiment 2 that were annotated by a k-means clustering algorithm searching for 3 clusters. Each sketch features strokes that form lines together, which the clustering algorithm was not capable of detecting.

Chapter 2

Learning a novel number system: The role of compositional rules & counting procedures

Sebastian Holt & David Barner

Department of Psychology, University of California, San Diego

Abstract

Humans count to indefinitely large numbers by recycling words from a finite list, and combining them using rules - e.g., combining sixty with unit labels to generate sixty-one, sixty-two, etc. Past experimental research has focused on children learning base-10 systems, and has reported that this rule learning process is highly protracted. This raises the possibility that rules are slow to emerge because they are not needed in order to represent smaller numbers (e.g., up to 20). Here, we investigated this possibility in adult learners by training them on a series of artificial number "languages" that manipulated the availability of rules, by varying the numerical base in each language. We found (1) that the size of a base - e.g., base-2 vs. base-5 - had little effect on learning, (2) that learners struggled to acquire multiplicative rules while they learned additive rules more easily, (3) that memory for number words was greater when they were taught as part of a sequential count list, but (4) that learning numbers as part of a rote list may impair the ability to map them to magnitudes.

Keywords: number; artificial languages; counting; syntax; base systems

Humans represent large, precise magnitudes, such as *two-thousand and five*, using culturally transmitted number words that are characterized by both memorized symbols and rules for combining them. Most number systems in use today feature a base-10 structure, in which words for 1-10 are combined to generate large numbers - e.g., by composing the unit labels for 1-9 with decade labels to generate number words up to 99, such as *sixty-six* (6*10+6). However, a much wider range of systems is attested on the historical record, including systems with bases as small as 2 and as large as 20 (Hammarström, 2010; Comrie, 2011; Table 2.1). For example, the Western Tribe of Torres Straits had a number system featuring only two lexical items, *urapun* (one) and *okosa* (two). These could be combined to create *okosa urapun* (three), *okosa okosa* (four), *okosa okosa urapun* (five), and *okosa okosa okosa* (six) (Haddon, 1890). These differences between historically attested systems are important because they may impact how number words are used, understood, and learned. Systems with small bases tend to differ in many ways from larger bases. For example, small bases may require less rote memorization because they feature only a few lexical roots, but may also require more frequent use of syntactic rules to construct larger numbers.[4] By contrast, systems with large bases may require more memorization of lexical roots, while depending less on rules to represent the same numbers. Although historically it was primarily adults who created and transmitted number systems, our current understanding of how they are learned has come overwhelmingly from studies of young children learning base-10 number systems. Consequently, it remains unclear whether our current understanding of how humans learn number words generalizes beyond base-10, where compositional rules are only relevant to learning relatively large numbers, which for children are only encountered relatively late in development. In the present study, we examined how the structure of number systems might impact learning by

[4] Importantly, combinations like *okosa okosa okosa* can be represented by a rule, or simply memorized, just as English words up to 100 can be rote learned or generated by rules.

investigating adult learners, who unlike children are capable of quickly learning new symbolic systems, and by training these learners on diverse number systems, with different bases and systems of rules.

Table 2.1
Several base systems: Western Torres (Haddon, 1890), Ventureño (Beeler, 1964; 1986), Kómnzo (Döhler, 2018), and English.

Language	**Base**	**Example Roots**	**Example Composed Numbers**
Torres	2	1, *urapun;* 2, *okosa*	3, *okosa-urapun;* 4, *okosa-okosa*
Ventureño	4	1, *pake'et,* 2, *'iškoṁ,* 3, *masəx,* 4, *tskumu*	6, *yəti'iškoṁ;* 12, *masəx tskumu*
Kómnzo	6	1, *näbi,* 2, *eda;* 3, *etha*; 4, *asar*; 5, *tabuthui;* 6: *nibo* [5]	7, *nibo a näbi,* 8, *nibo a eda …*
English	10	1, *one;* 2, *two;* … 10, *ten* [6]	16, *sixteen*; 60, *sixty*

How Are Number Words Usually Learned? Studies of number word learning in base-10 systems ranging from English, to Cantonese, to Hindi find that children acquire number words in a protracted process that takes 3-4 years, first learning meanings for small words like "one", "two", and "three", then learning to accurately count large sets. While children initially appear to treat the count routine as a meaningless, rote procedure, they ultimately learn rules for combining number words by around 5-and-a-half to 6 years of age (Almoammer et al., 2013; Barner et al., 2009; Carey & Barner, 2019; Ceylan & Aslan, 2018; Condry & Spelke, 2008; Davidson et al., 2012; Jara-Ettinger et al., 2017; Le Corre et al., 2016; Le Corre & Carey, 2007; Marchand et al., 2022;

[5] There are also roots expressing exponential powers of six, such as *fta* (36) or *taruba* (216).
[6] English also expresses exponents with new, non-composed roots, such as *hundred* (100) or *thousand* (1000). Tellingly, while higher exponents are composed, this composition is productive only in erudite speech (e.g., *billion*, *trillion*, *quadrillion*, etc).

Meyer et al., 2020; Negen & Sarnecka, 2012; Nikoloska, 2009; Piantadosi et al., 2014; Sarnecka et al., 2007; Sarnecka & Carey, 2008; Spaepen et al., 2018; Wagner et al., 2015, 2019; Wege et al., 2022; Wynn, 1990, 1992). This change occurs at around the same time that children begin to show an appreciation for how counting relates to number - e.g., that moving one place in the count list entails increasing the cardinality of the counted set by +1 (Davidson et al., 2012; Spaepen et al., 2018). Evidence that these older children use rules to produce larger number words comes from studies that ask them to count as high as they can, noting where errors occur. If children have memorized their count list without the use of rules, then errors should occur at random points - e.g., at 34, 47, etc. However, if they use a rule to compose the words for "one" through "nine" with familiar decade labels, errors should be clustered at decade transitions where children must either memorize the decade label (e.g., "sixty") or must have knowledge of a relatively sophisticated multiplicative rule (e.g., six-ty equals six-tens). Using this metric, previous studies report that while some children count up to random numbers, the most frequent error of 4- and 5-year-olds is to count up to decade transitions like 29, 39, or 49 (Fuson et al., 1982; Gould, 2017; Schneider et al., 2020, 2021a, 2021b; Siegler & Robinson, 1982; Wright, 1994). Importantly, these children, unlike those who count to random numbers, are often able to continue counting upwards when prompted with the next number - e.g., counting up to 49 when told that the number after 39 is "forty" - consistent with the use of an additive rule.

Because previous work has focused exclusively on children learning base-10 counting systems, it remains unclear whether this protracted learning sequence generalizes to other possible number systems, or whether rules might emerge earlier given other systems, such as those with smaller bases. One possibility is that children exposed to a base-10 system struggle to learn rules because of conceptual limitations, or because these rules depend on knowledge acquired at earlier stages of number word learning. However, it's also possible that rule learning emerges late not

because it is especially difficult, but because rule structures are not frequent within the small range of numbers that children hear in their language input and actually use. In English, the basis for many conclusions in the child literature, numbers up to 20 must be memorized, and rules are only salient in numbers greater than 30 (the first point at which the 1-9 rule is repeated). This is important because corpus studies find that although children are likely to hear small number words with high frequency in caregiver speech, larger words that exhibit rules are much less frequent (Willits, Jones, & Landy, 2016). Relatedly, number words are often taught to children as part of a memorized counting routine, which may encourage rote memorization over rule learning (Fuson, 2012). Meanwhile, historically humans have used a much more diverse set of number systems, with both smaller and larger bases, and often without the use of memorized counting systems (Hammarström, 2010; Comrie, 2011). This raises the question of whether different number systems, used in alternative ways, might lead to easier learning of numerical rule structures.

Some previous studies have begun to address how learners represent the syntactic rules of base systems using experimental methods (Barrouillet et al., 2010; Cheung et al., 2016). However, these studies have not directly tested how variability in the structure of number systems - like that found on the historical record - might impact learning. For example, Cheung et al. (2016) taught English- and Cantonese-speaking 4- to 6-year-olds a novel number term, *gobi*, referring to groups of three. They found that children speaking both languages could generalize a digit-multiplier construction (e.g., "one *gobi* houses" = 3 houses) to identify sets of six ("two *gobi* houses"), while only Cantonese-speaking children frequently succeeded in generalizing a bare form (e.g., "*gobi* houses" to "two *gobi* houses"). While there may be many linguistic or cultural reasons for this different performance (Cantonese's more transparent compositional syntax, its use of classifiers, etc.), this study demonstrated that children speaking both languages can learn a compositional rule that is central to a base-3 system. Barrouillet, Thevenot, and Fayol (2010) used a serial recall task

to test whether children have implicit representations of number syntax, even if they are not yet capable of consciously using those representations. Reasoning that implicit representations of syntactic structure in long-term memory should facilitate short-term memory storage and recall (e.g., Hulme, Maughan, & Brown, 1991; Gathercole, 1998), they asked 5-year-olds to repeat sequences of number words that were either grammatically well-formed (e.g., "three hundred forty five") or ill-formed ("five three forty hundred"). They found that children were more accurate when repeating well-formed number words, suggesting that they already had an implicit representation of the valid rules of composition. Although these studies suggest that children are sensitive to existing rules and can learn new ones, neither addressed how learning might differ across different systems, and whether it is impacted by factors such as the size of the memorized base, or the types of compositional rules present in a system (e.g., additive, multiplicative, exponential).

Does Counting Facilitate Rule Learning? Aside from leaving open how diversity in number systems impacts learning, previous studies also leave open how rule learning is impacted by the way in which we teach number words. As already noted, when children learn number words, it is typical for caregivers and early educators to explicitly train them on both the count list and related counting procedures. Such training typically includes learning to recite a subset of the count list in order (e.g., labels for 1-10 or ultimately 1-100), and the practice of placing these labels in one-to-one correspondence with objects in an array (e.g., pointing at each object while reciting the number words; Blevins-Knabe & Musun-Miller, 1996; Gibson et al., 2020; LeFevre et al., 2009; Turan & De Smedt, 2022). Notably, this is not how children learn other words in language. For example, although adult learners of a second language may memorize morphological paradigms, children learning their first language generally acquire rules by hearing different forms used in conversation

- e.g., abstracting the past tense or pluralization rules from hearing many instances of regular forms in caregiver speech, not by rote memorizing a verbal paradigm within an ordered sequence (e.g., *help, helped, helping*, etc.). Also notable is that not all attested number systems in the historical record were accompanied by counting practices. For example, in the number system of the Western Tribe of Torres Straits, described above, expressions *urapun* (one), *okosa* (two), and *okosa urapun* (three), were not memorized in a list and taught in scripted routines and procedures, but instead were learned like other morphological forms of the language, akin to English quantity expressions like "a couple", "several", and "a few" (Haddon, 1890). Similarly, although children learning languages like Slovenian and Arabic are taught counting routines and procedures, they are not taught routines for learning singular and dual morphology in their languages, although these also encode number (Almoammer et al., 2013; Marušič et al., 2016, 2021). Instead, these forms and the sometimes complex rules they entail are learned from exposure in conversation.

While there is evidence that early mastery of counting may have long-term educational benefits (Koponen et al., 2019; Martin et al., 2014), we know of no study that tests whether training on rote counting routines facilitates learning above and beyond frequent exposure to number words in conversational speech, nor how different types of training impact rule learning. One possibility is that rote training enhances a learner's ability to discover the rules that govern counting, by providing them with a robustly represented sequence that they can then analyze and decompose. That is, memorizing a count sequence may provide a dataset that supports further learning. Alternatively, it's possible that training with counting routines has a negligible or even negative effect on rule learning. For example, if the main impetus for constructing rules is to generate novel forms, memorization may supplant this need, by providing an alternative representational format. If learners receive sufficient rote training on numbers 1-100, for example, they may not need to construct rules that generate these forms. On this possibility, learners who are exposed to numbers

in both everyday language use and a memorized routine may actually be less able to generate novel forms vs. learners who are exposed to the forms only in language use.

Why Study Adults? Although almost all previous work on number word learning has been conducted in child learners, in the present study we began the investigation of these questions by testing fully numerate adults. We did so as a first step in this research program for two main reasons. First, whereas we know that children take years to learn even labels for 1, 2, and 3 and are difficult to train on even a single new number word (Carey et al., 2017; Huang et al., 2010; Spaepen et al., 2018), adults can more easily learn a large number of novel words in a single session. Also, testing adults allows us to dissociate the problem of learning a combinatorial rule system from the various other factors that may limit learning in children, including both domain general limits, like working memory and attention, but also factors specific to number, such as the fact that number words are almost always trained as part of a counting routine. By testing adults, we were able to train learners on complete number systems featuring multiple rules and lexical items, while also manipulating whether words were presented as part of a sequential counting routine. Finally, testing adults allowed us to administer multiple tasks, including tests of recall, generalization to untrained numbers, and semantic understanding of novel combinations.

Second, we know little about how adult learners learn novel number systems despite the fact that, historically, it was primarily adults who created and used number systems. This is important both because adult learners have greater cognitive capacities that may lead to the creation of different rule structures, but also because adult learners bring additional conceptual understanding to the problem of number word learning. In fact, there is evidence that some attested number systems emerged historically in communities that used other, pre-existing number systems. For example, the Torres Straits system existed in parallel with a separate, non-

compositional counting routine (Haddon, 1890). A similar pattern is found in the current day Nadahup languages, where some dialects are restricted to small numbers, some have a small numbers and a rote body counting system, and other dialects extend this body counting practice into a system of large, compositional, verbal numerals (Epps, 2006). In multilingual contexts existing number systems are often supplemented by imported systems, resulting in compound number systems with distinct linguistic sources (e.g., Garifuna, Mundurukú, Tsimané), or in which multiple number systems are used side-by-side, often for different purposes (e.g., Korean, Kómnzo). Finally, early written numerals (e.g., in Mesopotamia) were slow to develop, but almost certainly did so in the context of existing verbal number systems (Overmann, 2018). These historical observations suggest that the evolution of number systems often occurs in contexts of adult "multi-numeracy" (Marques, 2022), through the interaction of already-familiar number symbols with new symbols that are being constructed.

Testing adult learners allowed us to ask a series of novel questions. For example, we reasoned that if systems with smaller bases are harder to learn than systems with larger bases, then this difficulty should be reflected in the learning behaviors of adult participants. Similarly, if additive rules (e.g., adding unit labels for 1-9 to decades) are easier to learn than multiplicative rules (e.g., multiplying units by a base), then we expect that this should be reflected in the data of adults. Although it is possible that prior exposure to a base-10 system might make it easier for adult learners to acquire a new base-10 system, we did not expect such experience to transfer to learning systems with smaller (or larger) bases, since the additive and multiplicative rules in different base systems result in dramatically different structures for the same finite range of numbers. To make this concrete, consider some examples of base-2, 4, or 5 number systems that participants might encounter in the studies we describe here (see examples 1-4). Like English, these languages use rules to combine numbers additively or multiplicatively, expressing

exponential powers of the base using a new basic word (e.g., English 10^2 is *hundred*, 10^3 is *thousand*). A base-2 system, such as in (1), expresses 'three' compositionally as *2+1*. By contrast, in a base-4 or base-5 system, as in (2) or (3), it has a dedicated syllable.

(1) Base-2

Root Lexemes: (1) ka; (2) si; (4) tu[7]; (8) he

Numbers 1-10: ka, si, sika, tu, tuka, tusi, tusika, he, heka, hesi

(2) Base-4

Root Lexemes: (1) ho; (2) na; (3) se; (4) ki

Numbers 1-10: ho, na, se, ki, kiho, kina, kise, naki, nakiho, nakina

(3) Base-5

Root Lexemes: (1) ti; (2) ku; (3) so; (4) ne; (5) ha

Numbers 1-10: ti, ku, so, ne, ha, hati, haku, haso, hane, kuha

Notably, systems with different bases vary with respect to the number of root lexemes that a learner must memorize, the number of words that are composite and involve rules, and also the frequency of more complex rules. For example, in a base-10 system, the numbers 11-19 can be generated by a purely additive rule, by composing the lexical label for 10 (e.g., *zu*) with the labels for 1-9 (e.g., *ka, do, go,* etc.) to derive a composite numeral (e.g., *zu-ka, zu-do, zu-go,* etc.). By contrast, a base-3 system requires learners only to memorize 3 basic unit labels and 1 exponential label (for 9), but to compose numbers between 11-19 using both multiplicative and additive rules. For example, 16 might be expressed as 9+2*3+1. Nonetheless, learners of either system might be inclined to rote memorize numbers in their training set – compositional or not – and might remain

[7] N.B., these are exponential labels.

unaware of the generative rule structure governing each system. An important question is whether learners in fact notice and exploit the rule-governed structure of smaller base systems, and if so, which properties they learn and how they make use of these rules.

The present study. In Experiment 1, we investigated a small sample of possible number systems that varied in numerical base, which therefore also varied in many of the properties described above, such as the prevalence of additive rules or size of lexicon. In the task, participants learned a number system exhibiting one of six different numerical bases (2, 3, 4, 5, 8, 10). We began by pairing the Arabic numeral "1" with a randomly generated mono-syllabic novel number word, e.g., "ka", and then immediately tested the participant by asking them to type the name for the number 1 in this new language. Next, they saw the labels for 1 and 2 presented in sequence with their Arabic numeral translations, and then were tested on both. This progressive introduction of numbers continued up to the number 10 akin to a counting routine, with the result that smaller number words were trained more frequently than larger ones, resulting in frequencies loosely corresponding to those in natural language (e.g., BNC Consortium, 2007). This Training Phase exposed participants to a memorized list, but also allowed us to test memory for novel items, and whether memory differed across the different base sizes, and rule structures. However, the Training Phase did not necessarily require learners to notice or acquire the rules that characterized the different systems, since they could choose to simply memorize the sequences, without further analysis. For this reason, we also administered a Generalization Phase, in which we asked participants to generate numbers beyond their training set (e.g., 11, 12). Here, we expected that the

relative difficulty of noticing - and then deploying - different rule types might be more apparent in participants' responses, even if it is not obvious from errors in the Training Phase data.[8]

In Experiment 2, we explored two additional questions. First, whereas learners in Experiment 1 were always exposed to numbers in an ordered sequence, analogous to how children are often taught to count, in Experiment 2 we sought to isolate the effect of counting on learning from the effect of being exposed to a compositional rule structure. To do so, we trained participants in three distinct conditions, in which they learned number systems with (1) compositional syntax and a consistently ordered count list as in Experiment 1, (2) compositional rules but no consistent order of exposure, or (3) *no* compositional rules, but a consistently ordered count list. This allowed us to ask whether learning a system in the context of an ordered count routine impacted memory for the training set, but also whether it affected the ability to generate numbers beyond the training set. One possibility, for example, is that adding a count sequence to training improves memory for items and consequently improves the ability to extract rules and generate larger numbers. Alternatively, it may be that adding a count sequence to training supplants the need to extract rules. If so, learners might have good memory of items in the training set, but an impaired ability to generate larger numbers. Finally, in addition to testing these questions, we added a semantic task to Experiment 2, and asked participants to make quantity comparison judgments for trained numbers (i.e., judging which of two numerals represents a larger cardinality). This allowed us to probe whether the properties of different number systems impacted only the learning of grammatical forms, or also impacted the ability to reason about the semantic representations of these forms.

[8] Note that, given our interest in testing generalization, it was not possible to test larger base systems (e.g., base-20) since this would require the task to have been roughly four times as long as it was. We therefore limited ourselves to bases up to 10 for the current experiments.

Experiment 1

In Experiment 1, participants were assigned to learn one of six novel number systems that differed with respect to the size of their numerical base, and therefore their rule structures. In particular, in Experiment 1a we tested bases 2, 3, 4, and 5, and in Experiment 1b we tested the bases 8 and 10 in order to explore a more diverse range of memorized vs. rule-based number words. Participants first received a Training Phase in which we taught them labels for numbers in sequence and tested their memory for these labels, and then a Generalization Phase, in which we tested their ability to generate novel labels for numbers up to 15 in Experiment 1a, and 20 in Experiment 1b. Using these methods, we asked (1) whether some of these numerical bases are more or less learnable than others, given the relative difficulty of learning memorized words and rules, and (2) how participants generate number words, and whether some rules are easier to extract and deploy than others. For example, if additive rules are easier to learn and use than multiplicative rules, we may expect that learners of a base-8 system will find it easier to generate numbers up to 20 than learners of a base-4 system, which requires more frequent use of multiplicative rules.

Methods

Hypotheses and analyses for Experiments 1 and 2 were pre-registered at on the Open Science Framework at https://osf.io/rwqk7/ Materials and code for both experiments are available at: https://github.com/SebastianHolt/counting_syntax.

Participants

Based on a pre-registered power analysis, 235 adult participants were recruited online via Prolific, an online crowdsourcing platform commonly used for scientific research. Of these, 53 were excluded due to our pre-registered criteria, e.g., less than 50% accuracy during the Training Phase, accuracy 3 standard deviations below the mean, or navigating away from the experiment

screen on more than 5 trials. 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 (*mean* = $1.90). The most commonly reported first languages were English (*n*=87), Polish (*n*=23), Portuguese (*n*=21), and Spanish (*n*=13).

Stimuli

Number systems for each game combined lexical roots via a compositional syntax. Mono-syllabic roots were randomly generated from consonant-vowel bigrams (e.g., *ka*) such that no consonant or vowel was used in more than one root per number system. Rules for the different bases were constructed according to the following logic (see Table 2.2). For each base, *b*, (1) lexical roots denoted quantities up to *b*, (2) quantities larger than *b* but not divisible by it were represented by additive rules, where roots were added to base, (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 root (like *hundred*, 10^2), composed like the unit terms. Across the two experiments there were six bases: 2, 3, 4, 5, 8, and a base-10 system with 10 roots and no rules.

Table 2.2
Sample Base-3 Number System

	Base-3 Example	**English Example**
Additive	4 is *hi-no* (3+1)	*Fifteen* = 5 + 10
Multiplicative	6 is *ka-hi* (2*3)	*Twenty* = 2 * 10
Exponential	9 (3^2), *se*, is a lexical root	*Thousand* = 10^3
Roots (Base-3)	1 = *no*; 2 = *ka*; 3 = *hi*; 9 = *se*	
1-10 (Base-3)	*no*, *ka*, *hi*, *hino*, *hika*, *kahi*, *kahino*, *kahika*, *se*, *seno*	
Roots (Base-4)	1 = *no*; 2 = *ka*; 3 = *hi*; 4 = *se*	
1-10 (Base-4)	*no*, *ka*, *hi*, *se*, *seno*, *seka*, *sehi*, *kase*, *kaseno*, *kaseka*	

Note. The compositional rules of an example artificial base-3 number system, showing how 4 lexical roots are combined to label quantities up to 10, with reference to analogous rules of composition in English.

Training Phase

After reading an instructions page and passing a quiz that probed their comprehension of the instructions (see Supplementary Materials), participants entered the Training Phase, during which they learned a sequence of 10 novel number words (Figure 2.1). Each participant learned a sequence that was generated by 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, training resulted in the memorization of an ordered sequence, in which the frequency of exposure to each

number corresponded roughly to their frequency in natural language. At the end of this process, participants rehearsed the sequence one last time, from 1-10. Responses for these 64 recall trials (all responses in the Training Phase except the first exposure) were used to assess the relative learnability of the different base systems, in addition to data from the Generalization Phase. To motivate accurate responding, 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. Accuracy and response times were calculated only for the first try, and participants were shown the correct answer after responding incorrectly on all three tries. After submitting a correct response, or after their third try, participants were advanced to the next word in the sequence. They were instructed to complete trials "as quickly and accurately as possible", although they were not given a time limit.

Generalization Phase

After completing the Training Phase of the experiment, participants completed the Generalization Phase, wherein they were shown a table 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. They were then asked to enter number words from the language they had learned in the text entry fields, one at a time. Answers could not be edited after being entered. The last five numbers (i.e., 11-15 in Experiment 1a) or ten numbers (11-20 in Experiment 1b) required participants to generalize the number system to novel cardinalities that they had not yet encountered.

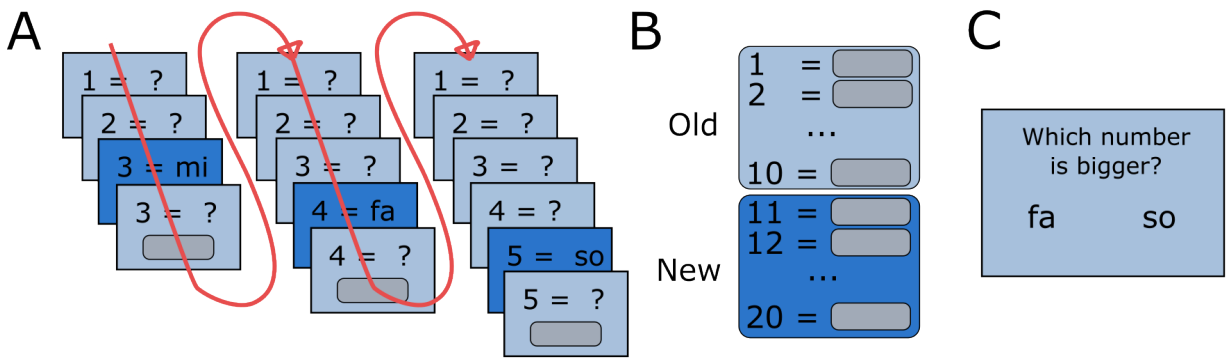


Figure 2.1. Experiment Phases. A. In the Training Phase, participants repeatedly typed a counting sequence, learning one additional number word with each repetition, until they had counted up to 10. B. In the Generalization Phase, participants typed the entire counting sequence, and then also labeled quantities from 11 to 15 (in Experiment 1a) or from 11 to 20 (Experiment 1b). The sequence of both old and new numbers appeared together on the same screen. C. Experiment 2 also contained a Magnitude Comparison task, in which participants clicked on the larger of every pair of number words they had trained on.

Results

Training Phase. To test whether base size impacted the learnability of a counting system, we assessed both accuracy and response time. To measure response time, we recorded the time that elapsed between receiving a prompt and the participant's first keystroke, and added this to the time they spent typing the response, divided by its character length.[9] Overall, accuracy during the Training Phase was near ceiling across conditions (.957, 95%-CI: [.953, .96]; see Figure 2.2), and the mean response time across conditions was 2.034 seconds (95%-CI: [2.002, 2.066]). Accordingly, mean accuracy did not vary much across base conditions (base-2: m = .96, sd = .04; base-3: m = .93, sd = .05; base-4: m = .98, sd = .05; base-5: m = .94, sd = .07; base-8: m = .97, sd = .03; base-10: m = .94, sd = .07), while response time varied more while showing similar qualitative trends (base-2: m = 1.74, sd = .96; base-3: m = 2.13, sd = 1.08; base-4: m = 2.12, sd = .96; base-5: m = 2.12, sd = .87; base-8: m = 1.75, sd = .72; base-10: m = 2.04, sd = .84). Thus,

[9] Note that response time analyses were repeated on reaction time – time from stimulus presentation until the participant started typing – to similar qualitative results.

overall participants performed with very high accuracy, resulting in few qualitative differences across conditions.

To assess these data statistically, we constructed separate linear mixed effects regressions predicting accuracy and response time from base-size, with random intercepts for each participant. We found no effect of base size on response time during the Training Phase in either Experiment 1a (bases 2-5: b=22.29 ms, t=.307, p=.759) or Experiment 1b (bases 8 & 10: b=-26.61, t=-.741, p=.460). Meanwhile, we found a negative effect of base size on accuracy in Experiment 1b, such that accuracy was poorer for the base-10 system relative to the base-8 system (b=-.223, z=-3.103, p=.002), though there was no such effect on accuracy in Experiment 1a (b=-.005, z=-.046, p=.963), or Experiment 1 as a whole (b=-.012, z=-.315, p=.753). At first pass, these results appear to suggest that these different systems are similarly easy to learn. However, as we note below, each system may be similarly easy if learners approach the task purely as one of rote memorization, since the same number of words was learned in each base system. If participants instead extract rules from training, then two measures may distinguish base systems from each other. First, the specific rules of composition within each base system - such as additive or multiplicative rules - may be more or less easy for participants to learn. Second, participants learning each base system may differ in their ability to generalize to new exemplars, where knowledge of rule structures is more important.

We therefore first asked whether different syntactic rules were easier or harder for participants in the Learning Phase. To do so, we conducted a *post hoc* analysis of data from the Training Phase to test whether participants' responses were faster or more accurate when the number words they learned contained more (or less) compositional rules. We coded each number word for the presence of an additive rule, a multiplicative rule, or a lexical root that represented an exponential meaning (i.e., 4 in the base-2 condition and 9 in base-3). These codes were then entered as predictors in a regression that also included each word's length in syllables, as well as which

base system it belonged to. This analysis found that the presence of multiplicative rules, but not additive rules, was 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). The greater difficulty of expressions containing multiplicative rules suggests that those words were harder to learn than words composed from additive rules (Figure 2.3), and therefore that at least also some learners were impacted by the presence of rules (whether they were aware of them or not).[10] At the suggestion of an anonymous reviewer, we conducted an additional *post hoc* analysis in which we restricted data to only those pairs of expressions in which one member of the pair was composed of two numbers added together (e.g., in base-3, the number 5 was composed as 3+2), and the other member was composed of those same numbers multiplied together (e.g., 6 is 2*3). This allowed for a more tightly controlled test of additive vs. multiplicative rules, as the only difference between each minimal pair was the additive or multiplicative rule itself. One such pair of numbers existed in each of the base-3 (3+2 and 2*3), base-4 (4+2 and 2*4), and base-5 (5+2 and 2*5) conditions, but not the other bases. We removed irrelevant predictors from the above linear models, leaving only base size (3, 4, or 5), and additive vs multiplicative rule type. Responses to the multiplicative forms were slower (b=652.599, t=4.462, p<0.001) though not significantly less accurate (b=-0.437, z=-1.636, p=0.102) than additive forms. This result is consistent with our primary analyses, which suggest that expressions containing multiplicative rules were more difficult to learn than those that contained only additive rules.

[10] While this suggests that bases with multiplicative rules should be harder, all else equal, this was only one of several features that varied between bases. For example, despite having the same number of multiplicative terms, base-4 system accuracy was greater than that of base-3 (b=1.539, z=4.137, p<0.005). One reason might be that the count list in the base-4 condition was easier to learn because each "decade" contains more exemplars of each rule type, providing more evidence to the learner for compositional rules. Alternatively, base-4 might be more amenable to rote learning, because the repetition of exemplars in sequence (e.g. 4-1, 4-2, 4-3) continues in longer, uninterrupted chunks.

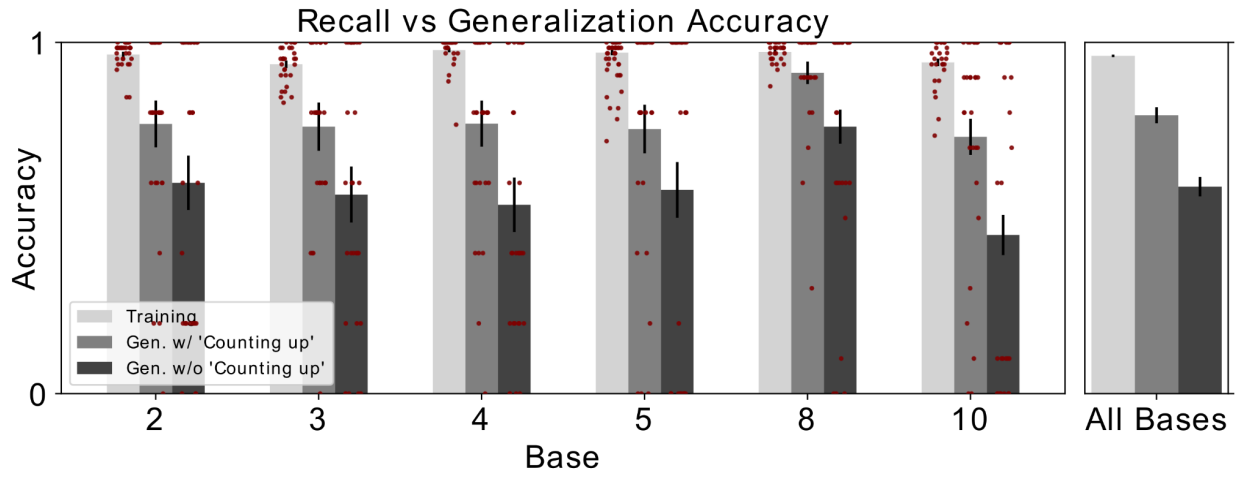


Figure 2.2. Counting in Recall and Generalization. Left Panel: Training Phase accuracy for recalling number words (light gray), compared to Generalization Phase accuracy permitting alternative rules or "counting up" strategy (medium gray), and Generalization Phase accuracy permitting alternative rules but not "counting up" (dark gray). Red dots are participant means. Error bars are 95% confidence intervals. Right Panel: Same data, collapsed across bases.

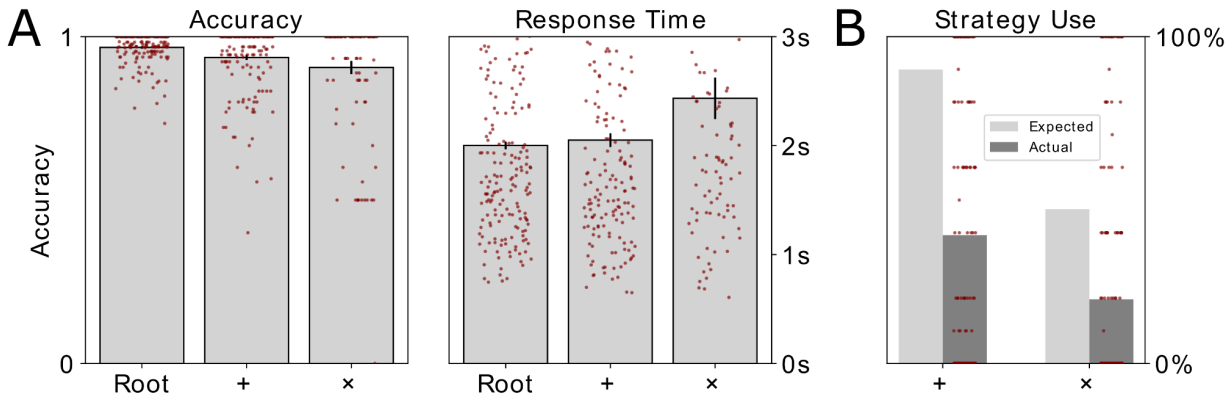


Figure 2.3. Use of Different Rule-Types. A. Word recall accuracy (left) and response time measures (right) in the Training Phase of Experiment 1, for number words featuring no compositional rules (i.e., lexical roots), additive rules, and multiplicative rules. Red dots are participant means. B. Strategy use in the Generalization Phase. Light gray indicates the proportion of number words that would be expected to show additive and multiplicative rules of composition for the extended range of numbers (11-15 for Experiment 1a, or 11-20 for 1b) under the rules used to generate each number system. Dark gray indicates the proportion of actual participant responses that could be parsed as using additive or multiplicative rules in that range.

Generalization Phase. Data from the Training Phase assessed participants' ability to recall words they had just learned, and provided only a preliminary test of how they mentally represented

the rules of each system. Although rules were available to be learned in most systems, learners could nevertheless ignore these rules and learn the number words by rote. To explore this question, we therefore analyzed data from the Generalization Phase.

Participants' Generalization Phase accuracy for Old Numbers (learned in Training Phase) was perfect (100%), confirming that they had learned the number system up to 10. Our first question was how many participants produced an "Exact Match" to target generalizations (i.e., forms that would be expected if participants had learned the rules that were used to generate their training set). For example, if a participant in the base-4 condition had learned *ho*, *se*, *na*, and *ki* to represent the numbers 1-4, then an exact match for the number 13 during generalization would be *na-ki-ho* (3*4+1). Of 182 participants across the different base conditions, 17 (9%) generalized the number system exactly in accordance with those rules through the entire generalization range (up to 15 in Experiment 1a, or up to 20 in 1b. Across participants, 37.7% of responses (454 out of 1205) were an exact match to the expected forms (CI: [0.349, .404]). This suggests that, despite being fully numerate users of a base-10 system featuring additive and multiplicative rules, most adult participants in our study did not learn the full set of additive, multiplicative, and exponential rules that were used to generate the number systems, even when trained in a base-10 system. A *post hoc* binomial linear regression with base size as a predictor, random intercepts for participants, and a random slope for number, revealed a significant effect of base size on the probability that participants would provide an exact match on each trial (b=0.435, z=2.057, p=0.04). This effect was driven by the base-8 condition and disappeared when base-8 data were removed from our analysis (b=0.162, z=0.848, p=0.396). This advantage for base-8 was likely due to the fact that, in this case, an alternative "counting up" strategy could generate the same responses as the target

rules for all numbers from 11 to 15. Having been explicitly taught 8+1 and 8+2, base-8 participants could continue this pattern until 8+7.[11]

Note that although most participants did not make Exact Match responses, this does not mean that most failed to use rules. Instead, participants may have constructed rules that deviated from those in their training set. We therefore asked whether participants may have constructed alternative rule systems. For example, if trained in the *ho, se, na, ki* base-4 system, participants may have created a purely additive expression for 13, such as *ki-ki-ki-ho* (4+4+4+1) or *na-na-se-ki-ho* (3+3+2+4+1), different from the target label but still mathematically coherent. They also may have created multiplicative expressions that differed from target responses, featuring both additive and multiplicative rules of composition, such as *se-ki-ki-ho* (2*4+4+1). We therefore coded participants' responses by first identifying substrings that were part of the number system they had learned (e.g., *ki* or *ki-se*) and then inferring rules of composition based on which parse produced the number closest to the target quantity. For example, while the response *se-ki-ki-ho* is ambiguous between a multiplicative (2*4+4+1 = 13) or purely additive parse (2+4+4+1 = 11), if it was produced as a label for the number 13, then we identified it as multiplicative because only this matched the target number. Because in natural language (and in the artificial language stimuli of this experiment), multiplicative rules of composition typically place the smaller number expression before the larger number (e.g., Mandarin *èrshí*, "two-ten", is 2*10) rather than vice-versa (e.g., *shíèr*, "ten-two", is 10+2), we only entertained multiplicative compositions for bigrams in which the first element denoted a smaller number than the second element.

Using this framework, we found that under a more Expanded Additive coding scheme, which included the Exact Match scheme as well as the above set of additive rules, participants

[11] That the "counting up" strategy did not explain as many responses in the base-10 condition, where it would also be consistent with the expected targets, might be explained by competition with those participants' bias to calque the Arabic numerals (see below).

produced responses that could be considered to be a correct product of rules on 53% of trials (CI: [.506, .562]), which was significantly greater than the 37.7% for the Exact Match scheme (χ^2(2, N = 1205 trials) = 120.405, $p < 0.001$). Meanwhile, we found that under the more permissive Expanded Multiplicative + Additive scheme participants produced correct responses 58.9% of the time (CI: [.561, .617]), significantly greater than under the Expanded Additive scheme (χ^2(2, 1205 trials) = 14.529, $p<0.001$), and therefore also greater than the Exact Match scheme.[12] To better understand this result, we considered how often we might expect participants' responses to fall into these categories by chance. For every label generated by a participant, we asked how likely it was that a random string of the same length would have been marked as correct under the same coding scheme. For example, if a base-4 participant's label for 11 was formed by combining "2*4+3", this was coded as consistent with the Expanded Multiplicative + Additive scheme. To assess chance, we asked how many randomly generated trigrams from the base-4 lexicon could also be interpreted as 11, using additive or multiplicative rules. For responses marked under Expanded Additive, we computed chance to be 6.9% on average (highest in base-3 at 11.1%, lowest in base-10 at 5.2%), and 8.4% for Expanded Additive+Multiplicative (highest in base-3 at 10.7%, lowest in base-10 at 6.3%). These rates thus tended to be much lower than the number of responses assigned to each strategy within our coding scheme.

Although base-10 knowledge appeared to be of little help in learning the target rules of novel base systems, it may nevertheless have influenced how learners generalized when using these alternative rule systems. For example, in the "Expanded Additive" and "Counting Up" strategies, it is possible that some learners may have directly calqued Arabic numerals by

[12] Consistency of rule use varied by participants, bases, and rule types. Using 80% consistency as a threshold, the following number of participants were consistent: Additive rules: base-2 (16/31), base-3 (9/30), base-4 (0/32), base-5 (0/30), base-8 (19/30), base-10 (8/29); Multiplicative rules: base-2 (0/31), base-3 (1/30), base-4 (11/32), base-5 (18/30), base-8 (0/30), base-10 (3/29).

associating each digit with a number word. In a base-2 system, for example, a learner might express 13 using the words for "one" (a lexical root) and "three" (a composition of 2-1), resulting in an expression glossed as "1-2-1". This occurred quite rarely in base-2 (3.2%, 95%-CI: [0.4, 6.0]), base-3 (5.3%, CI: [1.7, 8.9]), base-4 (3.1%, CI: [0.4, 5.8]), base-5 (4.0%, CI: [0.9, 7.1]), and base-8 (0.0%, CI: [0.0, 0.0]), though was somewhat common in base-10 (29.7%, CI: [24.4, 34.9]). Another influence from base-10 may have been responses that combined the expression for 10 (e.g. "8-2" in the base-2 system) with unit words, e.g., expressing 13 in the base-2 system as "(8-2)-(2-1)", i.e., "10-3". This occurred somewhat more often in base-2 (31.6%, CI: [24.3, 39.0]), but less in base-3 (16.7%, CI: [10.7, 22.7]), and less still in base-4 (10.0%, CI: [5.3, 14.7]), base-5 (0.0%, CI: [0.0, 0.0]), base-8 (4.0%, CI: [1.8, 6.2]), and base-10 (0.0%, CI: [0.0, 0.0]). Thus, we found that learners sometimes adopted base-10 strategies, and that when they did both the prevalence and type of base-10 strategy differed across systems, and occurred most frequently but in different ways for the smallest and largest bases.

Our next question was whether participants who employed alternative strategies were more or less likely to over-rely on additive rules (as opposed to multiplicative rules) relative to learners who exactly matched the target rules. To examine this, we compared the frequency of these rule types under the two expanded coding schemes just described to their prescribed frequency in the Exact Match scheme. For example, according to the Exact Match scheme, in the base-8 system 9 numbers between 11 and 20 feature an additive rule (i.e., 11-15, and 17-20), while 5 numbers feature a multiplicative rule (i.e., 16-20). Relative to an Exact Match scenario, we found that participants showed a stronger bias to use more additive rules than multiplicative rules for creating new number words ($\chi^2(3, N = 1205 \text{ trials}) = 143.625$, $p < 0.001$). This result is compatible with the earlier conclusion that additive rules may be easier to learn than multiplicative ones, and also that they may be easier to deploy in generalization.

In addition to considering these alternative additive and multiplicative systems, we also asked whether participants might have used a simpler alternative that involved counting up. For example, consider a base-4 participant who generated the label *se-ki-ho* (2*4+1) to represent 9. Remembering that in the counting sequence, *ho* is followed by *se*, this participant might represent 10 by "counting up" and replacing the syllable *ho* with *se* – e.g., *se-ki-se* (2*4+2). Although such a heuristic would produce responses consistent with the Exact Match scheme in most cases, it may also sometimes generate incorrect answers. For example, instead of correctly counting up from *ho* in the previous example, a participant might instead count up from *se,* resulting in the expression *na-ki-ho* to represent 10 (which in fact adds to 8). In order to explore alternative rule uses, we coded cases like this – i.e., counting up from any component of a complex expression – as correct responses in the Counting Up coding scheme. Also, although in the example just provided *na-ki-ho* was produced by counting up from a correct response of *se-ki-ho*, we also considered cases where participants counted up from incorrect responses. So, if they replied *ho-ho* (1+1) when prompted with 4 and then provided *ho-se* for 5, we also coded this as correct. Under this scheme, accuracy was 79.3% (CI: [.77, .815]), which was higher than in the Expanded Multiplicative + Additive scheme ($\chi^2(2, N = 1205 \text{ trials}) = 302.953$, $p < 0.001$) and therefore also higher than in the Additive and Exact Match schemes. This finding suggests that at least some participants created novel numbers using a simple counting heuristic, rather than additive or multiplicative rules of composition.

Discussion

In Experiment 1, we found five main results. First, we found that participants readily recalled the number words that they were trained on, with no important differences between small and large bases. Second, we found that participants were both faster and more accurate at recalling number words in their training set when they were composed of additive rules vs. multiplicative

rules. Third, we found that when participants were asked to produce numbers beyond their training set, they produced forms that exactly matched the target rules of their training set on just over 1/3 of trials, though only 9% of participants did so consistently. While we found a *post hoc* effect of base size in the generalization phase, this appeared to be driven by participants in the base-8 condition, where *ad hoc* rules, like counting up, mirrored target rules. Fourth, compatible with this last finding, we found that participants often created their own, alternative, rules to generalize beyond their training set. They created labels consistent with the use of an expanded set of additive rules on just over 50% of trials, and consistent with a combination of multiplicative and additive rules on nearly 60% of trials. When we also coded "counting up" responses as correct, nearly 80% of all participant responses reflected some kind of rule structure. Finally, consistent with the finding that expressions with additive rules may have been easier for participants to learn, when participants created novel rules, these were more likely than expected to be additive than multiplicative.

Overall, these results suggest that participants were very good at rote learning of diverse counting systems, that many learned target rules, and that most participants imposed some kind of rule-like structure on their input when asked to generalize. Many used some form of compositional system, while others used a simpler counting-up strategy. Finally, despite being fully numerate, most participants learned systems that did not directly reflect existing base-10 knowledge. When they did so, the kinds of base-10 knowledge they applied to this problem also differed between bases.

In Experiment 1, counting systems were always presented in a sequential order, akin to a counting routine, together with a compositional rule structure. In the Introduction, we noted that the presence of a counting routine might enhance the ability to rote memorize a number sequence, potentially reducing the need to derive rule structures. Given this, and given the finding that many

participants sometimes used counting-up in lieu of compositional rules, a natural question is whether there might be competition between these learning strategies, and whether learning a number system that is embedded in a sequential counting procedure might inhibit the acquisition of compositional rules. To explore this question, in Experiment 2 we asked how learning number words as part of an ordered counting sequence impacts learning. In particular, we asked whether counting routines facilitate learning or if they may instead inhibit the ability to extract compositional rules.

Experiment 2

In Experiment 1, we investigated the learnability of different base systems, all of which were trained as part of a sequence that was ordered in terms of magnitude. On one hand, the presentation of number words in an ordered counting sequence may be important to learning the meanings of number words, since their meanings are defined by their ordinal relations to one another. On the other hand, 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 to extracting rules for creating new numbers. Instead, learners may be able to learn these rules just as they acquire many other grammatical rules, via cross-situational learning that includes many distinct tokens of words across different utterances (Akhtar & Montague, 1999; Fisher et al., 1994; Gleitman, 1990; Pinker, 1994; Aslin & Newport, 2012). Also, as noted in the Introduction - and suggested by the results of Experiment 1 - it's possible that the opportunity to memorize numbers as part of a sequence may even supplant the need to learn rules, therefore interfering with rule learning.

In Experiment 2, we explored this question by comparing three groups of participants presented with number systems featuring (1) compositional rules and a consistently ordered count list, (2) compositional rules but no consistent order of exposure, and (3) no compositional rules,

but a consistently ordered count list. As part of testing this question, Experiment 2 also included a magnitude comparison task in which participants were presented two words from the novel system and were asked to judge which word represented a larger amount. This allowed us to not only probe participants' acquisition of combinatorial rules, but also their acquisition of number word meanings, and whether this learning was impacted by exposure to compositional rules or a sequential counting procedure.

Methods

Experiment 2 was pre-registered on the Open Science Framework (for link, see Methods, Experiment 1).

Participants

We recruited 119 adult participants via Prolific, of whom 28 were excluded due to our pre-registered criteria, e.g., less than 50% accuracy during the Training Phase, or accuracy 3 standard deviations below the mean. Samples were N=30, 31, and 30 in the Order+Rules, Order-Only, and Rules-Only conditions, respectively. Compensation was targeted at $14/hour + performance bonus (*mean* = $2.36). The most commonly reported first languages were English (n=38), Polish (n=16), Portuguese (n=10), and Spanish (n=7).

Training Phase

Because Experiment 1 found no consistent effects of base size, all participants in Experiment 2 were taught a base-5 system, selected because it exhibits combinatorial structure over half the count list (from 1-10), and requires only additive and multiplicative rules (no exponents) to label any number up to 20 (the highest cardinality appearing in the experiment). In

order to independently manipulate the presence of combinatorial rules and an ordered counting procedure, we divided participants into three conditions. In Condition 1 (Order+Rules), number words featured compositional rules, and were trained as a consistent, ordered list, as in Experiment 1. 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. In other words, these combinations of syllables did not correspond to meaningful units (morphemes) and therefore were arbitrary, and could not be analyzed via rules. For example, if the lexicon learned in the Order+Rules condition was {1=*ni*, 2=*se*, 3=*ta*, 4=*ho*, 5=*ku*, 6=*ku-ni*, etc.}, the lexicon learned in the Order-Only condition might be {1=*ku*, 2=*ta*, 3=*ku-ni*, 4=*ni*, 5=*ho*, 6=*se*, etc.}. In Condition 3 (Rules-Only), the lexicon was identical to the Order+Rules condition, but the words were trained in a random order (i.e., not in ascending order of magnitude). Critically, the frequency of each number word in this condition was matched with the synonymous word in the Order+Rules 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 Experiment 1b, and tested numbers from 11-20.

Magnitude Comparison

After the Generalization Phase, participants completed the Magnitude Comparison task, in which they compared the magnitudes of number words they had learned. In each trial, they were presented with a pair of unequal number words from the numbers they had just learned, and asked to click on the one that represented the 'bigger' value (Figure 2.1C). 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 combinations of number words were queried.

Results

Training Phase. Accuracy during the Training Phase was comparable to Experiment 1 (.939, 95%-CI: [.932, .946]), as were mean response times 2.230 seconds (95%-CI: [2.167, 2.293]). Accuracy was generally high across all three conditions, though important differences existed, as discussed below (Rules+Order: m=.96 sd=.02; Rules-Only: m=.91, sd=.05; Order-Only: m=.95, sd=.02). Response times also exhibited some variability (Rules+Order: m=1.73, sd=.79; Rules-Only: m = 2.76, sd = .98; Order-Only: m = 2.20, sd=.99). To test the relative contributions of rules and order on recall of learned words, we constructed mixed effects linear regressions predicting 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 trained number words than in the Order+Rules condition, suggesting that learning an ordered count sequence helped participants remember words that they were explicitly trained on. Compatible with this, participants in the Order-Only condition were just as accurate as those in the Order+Rules condition (b=-.469, z=-.945, p=.345; Figure 2.4A), suggesting that the addition of compositional rules did not improve recall of trained words. Nonetheless, participants in the Order+Rules condition were slightly faster to respond than those in the Order-Only condition, suggesting that compositional rules may slightly facilitate recall (b=688.78, t=2.716, p=.008). This model also showed that participants were less accurate when responding to larger numbers (b=-0.141, z=-4.84, p<0.001), although they also made their responses more quickly (b=-44.16, t=-3.282, p=0.001).

While participants did not generally become more accurate with increased exposure (b=-0.011, z=-0.191, p=0.849), they did become faster to respond (b=-50.61, t=-2.257, p=0.024). With increased exposure, participants in the Order-Only condition were especially faster than in the Order+Rules condition (b=-78.74, t=-2.615, p=0.009) but not more accurate (b=0.114, z=1.414, p=0.158), whereas those in the Rules-Only condition were both faster (b=-130.13, t=-4.284, p<0.001) and more accurate (b=0.44, z=5.282, p<0.001). Together, these results suggest that - at least for numbers within a training set - learners benefited from learning number words in an ordered counting sequence and that the presence of transparent rules may have helped learners to recall those words faster, but not necessarily more accurately. While increased exposure (i.e., cumulative frequency) did not always predict higher accuracy, this may be due to the structure of trials in both the Order-Only and the Order+Rules conditions. Because the counting procedure increased in length each time it was repeated (i.e., 12, 123, 1234, etc.), successive trials of each number were also interrupted by a longer interval of trials for different numbers.

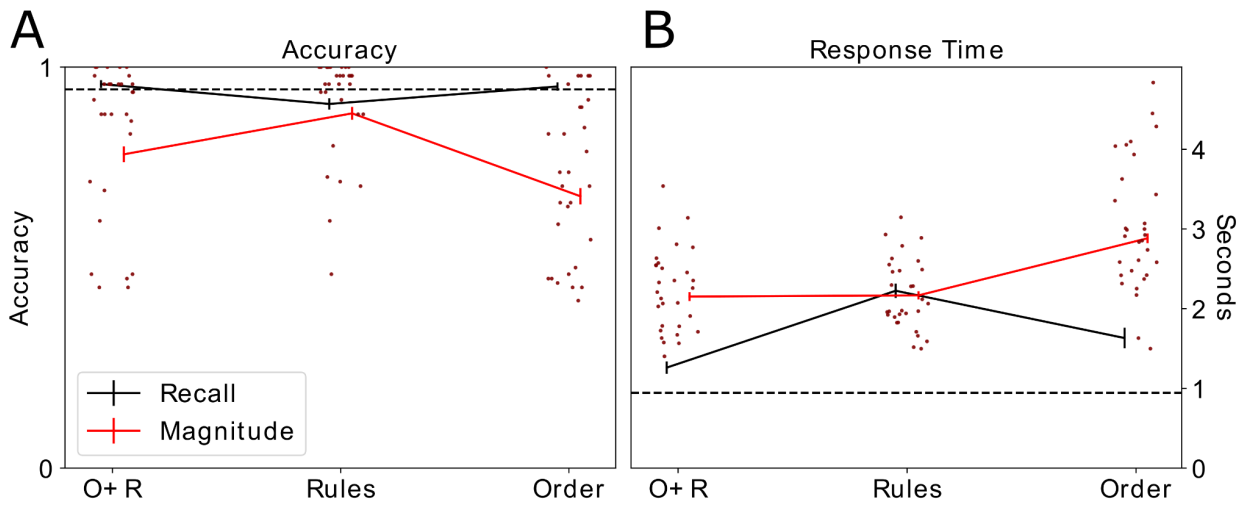


Figure 2.4. Counting and Rules in Number-Word Learning. A. When participants learn numbers in a counting procedure (O+R and Order), they recall words better (black) but make less accurate judgements of magnitude (red). B. Participants learning numbers in a stable counting order also recall number words faster (black), but do not judge the magnitude of numbers faster (red). Dotted line in each panel represents average of the base-5 condition from Experiment 1. Dots in both panels show mean Magnitude Comparison performance of individual participants.

Generalization Phase. Although data from the Training Phase address the impact of rules and order on learning trained items, they do not address how these factors impact the acquisition of rules that can be used to generate and comprehend larger numbers. Therefore, as in Experiment 1, we asked whether participants generalized the number system to the range 11-20 using expressions exactly matching our algorithm for producing artificial number words, whether they used other additive or multiplicative rules, and whether they showed signs of a counting up strategy. Because the system learned by participants in the Order-Only condition featured no compositional rules, these analyses could only be meaningfully applied to the Rules-Only and Order+Rules conditions. In these two conditions, exact matches to the target rules occurred on 35.8% of 600 trials (CI: [.32, .397]), comparable to Experiment 1. Participants' rate of success did not increase under an Expanded Additive coding scheme (.362, CI: [.323, .4]; $\chi^2(2, N = 600 \text{ trials}) = 0.029$, $p = .865$), but reached 53.3% under the Expanded Multiplicative + Additive scheme, which was significantly greater than under the Expanded Additive scheme (.533, CI: [.493, .573]; $\chi^2(2, N = 600 \text{ trials}) = 76.589$, $p < 0.001$). Accuracy was higher still, at 74.3%, when also allowing a counting up strategy (CI: [.708, .778]; $\chi^2(2, N = 600 \text{ trials}) = 138.687$, $p < 0.001$). Across these analyses, we found no effect of condition, suggesting that learning number words in a sequential order neither helped nor hindered participants' acquisition of compositional rules.[13]

Magnitude Comparison Task. The results presented thus far suggest that exposure to an ordered counting sequence helps learners to recall number words, but not to extract rules that generalize to novel forms. Left open by these analyses of how participants learn compositional

[13] In a pre-registered analysis, we also asked whether there was a difference in the diversity of forms created, measured using the Simpson's diversity index of forms given for each cardinality. We found there was no difference between the Rules-Only condition (.803, CI: [0.727, 0.879]) and the Order+Rules condition (.769, CI: [0.676, 0.862]).

rules is whether differences in training conditions impacted how the *meanings* of number words are acquired. To explore this question, we constructed mixed effects regressions to predict accuracy and response time on the Magnitude Comparison task, which required participants to reason about the magnitudes that novel number words represented. This included terms for condition (Rules-Only, Order+Rules, Order-Only), the sum of the compared numbers (a+b), the interaction between condition and sum, as well as intercepts for each participant.[14]

These models revealed that participants in the Order+Rules condition were faster (b=-1218.046, t=-6.481, $p<0.001$) and more accurate (b=2.019, z=4.171, $p<0.001$) than those in the Order-Only condition. Participants in the Rules-Only condition were even more accurate than those in the Order+Rules condition (b=2.586, z=4.299, $p<0.001$), though not significantly faster (b=-132.906, t=-0.692, $p=0.49$). Thus, these models thus showed a marked difference in performance between conditions, and that participants were generally less accurate when trained with an ordered counting sequence, while providing responses with equal speed. They also revealed that the magnitude of compared numbers impacted performance. In the Order+Rules condition, larger summed cardinality of the compared numbers predicted lower accuracy (b=-0.107, z=-6.203, $p<0.001$) and longer response times (b=13.292, t=2.048, $p=0.041$). This was also observed in the Rules-Only condition: lower accuracy (b=-0.233, z=-8.862, $p<0.001$) and longer response times (b=19.722, t=2.889, $p=0.004$). It was only in the Order-Only condition that larger summed cardinality predicted faster response times (b=-25.504, t=-4.004, $p<0.001$), though it did not predict significant differences in accuracy (b=-0.016, z=-1.133, $p=0.257$). Thus, overall it

[14] We had planned to include, as a predictor, the cumulative frequency of both numbers being compared. As observed above in reference to the Training Phase however, our ability to interpret the frequency predictor was limited by several elements of the task design. Most relevant here, frequency was heavily anti-correlated with the sum of the compared numbers (0.996), making the estimates for each predictor very sensitive to small changes in model parameters. We therefore removed frequency from the model.

seems that participants showed lower accuracy for larger numbers. However, our interpretation of the summed cardinality effect is limited by a non-linear relationship between cardinalities and word-lengths that existed only in the Order+Rules and Rules-Only conditions: pairs of small number words had the same word-length (i.e., one syllable), as did pairs of large number words (i.e., two syllables), but not pairs containing both a small and a large number, which also summed to cardinalities in between small-small and large-large pairs. We return to the question of word-length below.

The previous analyses indicated that participants made Magnitude Comparison judgments differently across conditions. We next asked whether these judgments were compatible with a numerical distance effect, in which magnitude judgments are faster and more accurate for numbers that are farther from each other on a number line than for numbers that are close to each other (e.g., 1 and 9 are easier to distinguish than 8 and 9). To do this, we constructed a set of linear mixed effects models predicting accuracy and response time from condition, the sum of numbers in each pair ($a+b$), and the numerical distance between them ($a-b$), with a random effect of participant. This model showed qualitatively similar estimates for the main effect of each condition, and a main effect of summed cardinality consistent with the average of the effects reported above – that is, no overall effect on reaction time (b=0.173, t=0.047, p=0.963), and a negative effect of summed cardinality on accuracy (b=-0.077, z=-7.88, p<0.001). It also revealed a main effect of numerical distance, in which larger differences between each pair of numbers resulted in faster responses (b=-106.55, t=-15.85, p<0.001) that were also more accurate (b=0.507, z=19.095, p<0.001).

Observing this, we asked whether this numerical distance effect differed between conditions, and if so, why. Previous studies (e.g., Krajcsi, Lengyel, Kojouharova, 2016) suggest that the sources of the numerical distance effect may differ between judgements of perceptual stimuli (e.g., dot arrays) and symbolic representations (e.g., Arabic numeral digits). The methods

used here differ from previous studies in two important ways, which we address in turn. First, the number words in two of the three conditions – Order+Rules and Order-Only – were learned in a counting sequence. We therefore asked whether the distance effect was impacted by the presence of this counting procedure. Second, the symbolic numbers compared in this task varied in length between one and two syllables, and when pairs of words differed in length, choosing the longer word was a reliable heuristic for choosing the larger number only in the Order+Rules and the Rules-Only condition. We therefore asked whether participants were more likely to show a numerical distance effect in the rules-based conditions, where the length of each number word being compared could also serve as a heuristic for magnitude. Finally, we asked whether the use of such a heuristic would fully account for any distance effect in these groups, or whether other characteristics of our artificial stimuli (e.g. being symbols, being counted) might also mediate the effect.

To explore these questions, we added an interaction term between numerical distance and task condition to the above model. This revealed that numerical distance predicted both higher accuracy and faster reaction time across conditions. We found greater accuracy for greater numerical distances for Order+Rules (b=0.825, z=12.42, p<0.001), Rules-Only (b=0.742, z=8.707, p<0.001), and Order-Only (b=0.353, z=11.293, p<0.001). And we found faster reaction times for greater numerical distances for Order+Rules (b=-149.11, t=-12.796, p<0.001), Rules-Only (b=-132.35, t=-10.813, p<0.001), and Order-Only (b=-47.925, t=-4.36, p<0.001). These results suggest that participants can learn to associate novel words with numerical magnitudes, both in the presence of syntactic rules or a counting procedure. Further, the numerical distance effect was strongest in the Order+Rules condition and not significantly weaker in the Rules-Only condition for accuracy (b=-0.083, z=-0.771, p=0.441) or response time (b=16.76, t=0.992, p=0.321), suggesting that a counting sequence did not contribute much to the numerical distance effect over

and above syntax. The numerical distance effect was also significantly weaker in the Order-Only condition compared to the Rules-Only condition for both accuracy (b=-0.388, z=-4.28, p<0.001) and response time (b=84.425, t=5.131, p<0.001), and therefore also weaker than the Order+Rules condition. This suggests that rules were more helpful than a counting procedure for learning the numerical content of novel number words.

We next asked what accounted for the numerical distance effect in the Order+Rules and Rules-Only conditions. In both conditions, many pairs of numbers could be successfully distinguished by comparing the length of words – any two syllable number word always encoded a larger value than a one-syllable number word. To assess this, we added one further predictor to the original model (condition + cardinality + distance), which was whether the pair of number words on each trial differed in length by 1 syllable, or contained the same number of syllables. Overall, we found that although word-length was a significant predictor of performance, it did not fully explain the numerical distance effect. Adding word length improved model fit for both accuracy (chisq = 102.35, p<0.001) and reaction time (chisq = 143.53, p<0.001), and predicted higher accuracy for both Order+Rules (b=2.412, z=7.142, p<0.001) and Rules-Only (b=3.716, z=5.027, p<0.001) conditions, with a smaller positive effect in the Order-Only condition (b=0.365, z=3.243, p=0.001).[15] Word-length also predicted faster reaction time in Order+Rules (b=-781.54, t=-10.893, p<0.001) and Rules-Only conditions (b=-811.69, t=-11.126, p<0.001), but not in the Order-Only condition (b=-4.81, t=-0.105, p=0.917). However, numerical distance by itself continued to predict higher accuracy in the Order-Only condition (b=0.348, z=11.151, p<0.001)

[15] This was likely because occasional congruence between word-length and magnitude in Order-only trials helped participants respond faster and more accurately than baseline, and not due to a word-length heuristic. Participants in the Order-only condition were as inaccurate comparing numbers of the same word length (.633, CI: [0.602, 0.663]) as comparing pairs that misleadingly differed in word length (.631, CI: [0.594, 0.668]), but performed much better when word length was congruent with magnitude (.863, CI: [0.897, 0.83]).

as well as faster response times (b=-47.92, t=-4.456, p<0.001). Numerical distance also continued to predict higher accuracy in the Order+Rules (b=0.502, z=6.219, p<0.001) and Rules-Only (b=0.338, z=3.076, p=0.002) conditions, but it no longer significantly predicted faster reaction times in the Order+Rules (b=-24.75, t=-1.534, p=0.125) or Rules-Only (b=-4.87, t=-0.294, p=0.769) conditions.

General Discussion

In two experiments, we trained numerate adults on a series of novel number systems and manipulated the availability of rule structures and ordered count sequences. This approach produced several important results. In Experiment 1 we found that base size had little effect on participants' ability to memorize number words in their training set. Also, generalization data from Experiment 1 showed that less than 10% of participants extracted rules matching those in their training, with little impact of base size on performance. While we found no overall effect of base size, we did find that associated features (like the presence of rules) did affect the learnability of each system. For example, generalization data also revealed that additive rules were easier to learn than multiplicative rules, and that participants often imposed novel structures onto their input, including additive, multiplicative, and counting-up strategies, in ways that varied with the base system they had learned. Thus, despite being numerate adults with prior exposure to both multiplicative and additive rules, most participants in our study struggled to learn the precise systems they were trained on, and instead hypothesized rules that were not contained within the training data. In Experiment 2 we found similar results, and also that learning number words in a consistently ordered count list helped participants to recall those words, but not to generalize the number system to describe novel quantities. Also, magnitude comparison data from Experiment 2 suggest that participants were more likely to learn the meanings of number words when trained on a system that featured only rules, compared to one that featured only a sequential counting

structure, or counting and rule structures combined. This suggests that presenting number words in a count routine may help learners memorize the words in their training set, but may also impair the ability to access quantity information.[16]

Effects of Base Size and Counting on Number Word Learning. In the Introduction, we noted that most previous studies of how learners acquire number systems have focused on how children learn base-10 systems that are presented in the context of counting routines. Consequently, little is known about the learnability of alternative, historically attested base systems, or how they are learned or used by adult learners. We also noted that understanding the learnability of different systems in adult learners might have consequences for how we understand results from the developmental literature. In particular, if systems with smaller bases are as easy to learn as larger bases, then it's possible that training children on systems with smaller bases would result in earlier learning trajectories, since smaller base systems expose learners to compositional rules more frequently. Also, because smaller bases contain fewer lexical items, they require less rote memorization before rules can be deployed. By contrast, if we were to find that smaller base systems are harder to learn for adults - e.g., because of their greater reliance on rules - we might conclude that training children on such systems would be unlikely to accelerate learning. Because our study did not test children directly, it remains to be seen whether these findings generalize to younger learners, owing to differences in their cognitive capacities and learning environments. Our results in adults, while not showing an advantage of smaller bases, also don't suggest that they are harder, though they may involve different learning processes. Given this, future studies should train younger learners with alternative base systems, and ask

[16] Note that participants could have used ordinal position to infer relative magnitude (e.g., as in a familiar number system), or may have sometimes translated from the novel system to English or Arabic numbers.

whether they are able to acquire combinatorial rules earlier in acquisition, if exposed to numbers that require their use.[17]

A second issue that we considered in the Introduction is how learning a counting routine might impact number word learning. Previous studies suggest that there may be long-term benefits of expertise with counting routines (Koponen et al., 2019; Martin et al., 2014), and it has been proposed that learners discover new conceptual content on the basis of the counting routine (Carey, 2004; 2009). However, no previous studies have tested whether training on rote routines facilitates learning of combinatorial number rules above and beyond frequent exposure to number words in conversation (though studies have tested how these impact earlier outcomes, like acquisition of small number words, or the cardinal principle; Mix et al., 2012; Gibson et al., 2020). Our study begins to address this question by testing adult learners, and suggests that embedding training in a count routine may benefit recall of trained words but not rule learning. Further, we found some evidence that embedding number words in a counting routine may even impede the acquisition of number word meanings. In particular, when participants were asked to decide which of two number words represented a larger magnitude, they performed significantly worse if they were trained with a sequential counting routine – with or without combinatorial rules – versus a system with just combinatorial rules alone. These results comport with a growing body of evidence that numerical reasoning is often closely tied to the structure of external representations of number. For example, judgments of magnitude are made differently when comparing symbolic numbers or perceptual magnitudes (Krajcsi, Lengyel, & Kojouharova, 2016; 2018; Krajcsi, 2017). Similarly, research on the history of number highlights how the medium of a number system can impact computation.

[17] As we note below, we are currently conducting such studies. We do not include child data here because these studies require substantial modifications to methods and a more detailed review of related developmental studies, and would therefore produce an unwieldy report if combined with the current studies.

Such work highlights the effects of visual properties of number symbols on computation, such as size-ordering, in which longer expressions reliably denote larger cardinalities (Chrisomalis, 2020), the separate uses of abaci and Roman numerals for calculation and recording of quantities (Schlimm & Neth, 2008), and the role that instantiating numbers in different media may play in the systematization of number concepts over time constitutive role in the conceptual trajectory of number systems (Overmann, 2018; 2023).

The Use of Artificial Language Learning in the Study of Numerical Cognition. Beyond these specific questions that motivated our study, our results also suggest that studies of artificial language learning can provide a fruitful method for further exploration of numerical cognition. A large existing body of work makes empirical claims about the nature of language by teaching artificial languages to linguistically savvy adults (see Kirby et al., 2014, or Culbertson, 2023, for review), but few previous studies have taken a similar approach to number systems. Despite the fact that participants in our study were fully numerate adults with sophisticated number concepts, they exhibited remarkable variability in their learning strategies, suggesting that their expertise did not fully determine their approach to learning new systems. In fact, even when trained in a novel base-10 system participants fared no better than when trained with alternative bases. While a small number of participants learned all target rules, others learned only target additive rules, but not multiplicative rules, despite the existence of both additive and multiplicative rules in the English number system. Some participants' were appeared to directly adopt strategies inspired by their base-10 knowledge, though this was less common, and when they did their particular form of base-10 strategy differed according to the base system they were learning. Finally, many participants created untrained rule systems. When they did, these rules often resembled those of historically attested systems. The variability of strategies we observed confirms the value of experimental methods for assessing the foundations of compositional rules in number systems. Consequently,

our findings set the stage for asking whether similar learning strategies exist in children or innumerate adults (e.g., Gordon, 2004; Pica et al., 2004; Frank et al., 2008; Pitt et al., 2022). Drawing on the lessons of this study, in which multiplicative rules and exponential roots may have introduced complexity to learning, future studies might begin with simpler systems in order to target specific tradeoffs, like that between memorizing lexical roots and learning a specific rule. For example, it is possible that exposing learners to a purely additive system, where only lexical learning and rule frequency differ, might result in more obvious effects of base size.

Open Questions and Future Directions. Our study leaves open a number of interesting questions. First, it leaves open why participants found it easier to learn additive rules than multiplicative rules. One possible factor might be that additive rules were more frequent in participants' training, and also occur more frequently in their familiar base-10 system. Another possibility is that participants were more likely to intuit the logic of additive rules because they are conceptually more concrete. For example, for an additive composition like 4+2, sets of both 4 and 2 are present, whereas for a multiplicative composition like 2*4, only groups of 4 are present, and the 2 must be interpreted as quantifying the sets of 4 themselves, rather than the objects contained within each set. Interestingly, the emergence of multiplicative rules likely post-dates that of additive rules in the history of most languages (Hurford, 1987). On some accounts, multiplicative constructions may only be invented when the expressive limit of a simpler, purely additive system is reached (Damerow, 1996). As an example, Damerow notes that the largest number in many body tally systems is the last body part within the tally (e.g., Eipo *seselekyaba*, "right little finger", meaning 25). To express larger cardinalities, this number may receive a separate designation, *yupe ton*, "one time counted", from which multiplicative expressions like *betinye ton*, "two times counted" are subsequently coined. The conceptual content of each kind of expression is clearly

different: rather than the literal body parts within the tally system, the multiplicative expressions refer to the counting procedure itself.

Also left open by our study is how social processes impact 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 multigenerational transmission processes, 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., 2022).

In addition to assessing the learnability of artificial systems for adult learners, future studies should consider whether naive learners, who have not previously acquired a number system, might show different learning strategies. Naturally, while our studies provide insight into the factors that impact the learnability of different number systems by numerate adults, they do not tell us how these factors impact child learners, where many additional factors likely impact learning. For example, aside from not having prior exposure to a number system, children differ in working memory capacity, their understanding of the social function of number, and also in the practices that they associate with counting. For example, in our study, adults learned number words as syntactic objects that referred to cardinalities (indicated by Arabic numerals) whereas when children learn to count they associate individual words with objects in a count, which may defocus

attention to the relationship between combinatorial number symbols and their cardinal meanings. Studies now in progress are exploring semi-numerate children's ability to learn different base systems, using methods adapted from the experiments presented here. 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 by observing the acquisition of natural language alone.

In this study, we probed the impacts of counting structures and procedures on learning by training numerate adults on a novel number system. We found that participants could learn the words of a new system effectively, regardless of the size of the numeral base. We also found that the rote counting procedure in which these number systems were learned helped learners to memorize each number word without necessarily acquiring the compositional rules underlying them. Instead, participants showed a strong tendency to rely on additive rules, rather than multiplicative rules, to generate novel number words. Further, we found that when number words were learned in random order rather than within a counting procedure, participants performed worse at cued recall of those words, but were better able to access their meanings when comparing the magnitude of number-word pairs. We conclude that the counting procedure and syntactic structure of a number system not only facilitate different aspects of number-word learning, and that their interaction may not always be entirely constructive.

Acknowledgement

Chapter 2, in full, has been submitted for publication and may appear as Holt, S. & Barner, D. (2025). Learning a novel number system: The role of compositional rules & counting procedures. Under review at *Cognitive Science*. The dissertation author was the primary researcher and author of this paper.

References

Akhtar, N., & Montague, L. (1999). Early lexical acquisition: The role of cross-situational learning. *First Language*, *19*(57), 347–358. https://doi.org/10.1177/014272379901905703

Almoammer, A., Sullivan, J., Donlan, C., Marušič, F., Žaucer, R., O'Donnell, T., & Barner, D. (2013). Grammatical morphology as a source of early number word meanings. *Proceedings of the National Academy of Sciences*, *110*(46), 18448–18453. https://doi.org/10.1073/pnas.1313652110

Aslin, R. N., & Newport, E. L. (2012). Statistical learning: From acquiring specific items to forming general rules. *Current directions in psychological science, 21*(3), 170-176.

Barner, D. (2017). Language, procedures, and the non-perceptual origin of number word meanings. *Journal of Child Language*, *44*(3), 553–590.

Barner, D., Chow, K., & Yang, S.-J. (2009). Finding one's meaning: A test of the relation between quantifiers and integers in language development. *Cognitive Psychology*, *58*(2), 195–219.

Baroody, A. J., & Price, J. (1983). Brief Report: The Development of the Number-Word Sequence in the Counting of Three-Year-Olds. *Journal for Research in Mathematics Education*, *14*(5), 361–368.

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.

Beeler, M. S. (1964). Ventureño Numerals. *Studies in Californian Linguistics*, 13-18.

Beeler, M.S. (1986). Chumashan Numerals. In M. P. Closs (Ed.), *Native American Mathematics*. University of Texas Press.

Blevins-Knabe, B., & Musun-Miller, L. (1996). Number Use at Home by Children and Their Parents and Its Relationship to Early Mathematical Performance. *Early Development and Parenting*, *5*(1), 35–45. https://doi.org/10.1002/(SICI)1099-0917(199603)5:1<35::AID-EDP113>3.0.CO;2-0

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

Briars, D., & Siegler, R. S. (1984). A featural analysis of preschoolers' counting knowledge. *Developmental Psychology*, *20*(4), 607.

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

Carey, S. (2009). *The Origin of Concepts.* Oxford University Press.

Carey, S., & Barner, D. (2019). Ontogenetic origins of human integer representations. *Trends in Cognitive Sciences*, *23*(10), 823–835.

Carey, S., Shusterman, A., Haward, P., & Distefano, R. (2017). Do analog number representations underlie the meanings of young children's verbal numerals? *Cognition*, *168*,

243–255.

Ceylan, M., & Aslan, D. (2018). Cardinal Number Acquisition of Turkish Children. *Journal of Education and E-Learning Research*, *5*(4), 217–224.

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. https://doi.org/10.1016/j.cogpsych.2016.11.002

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. https://doi.org/10.1111/cogs.12875

Chrisomalis, S. (2020). *Reckonings: Numerals, cognition, and history*. MIT Press.

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

Condry, K. F., & Spelke, E. S. (2008). The development of language and abstract concepts: The case of natural number. *Journal of Experimental Psychology: General*, *137*(1), 22.

Culbertson, J. (2023). Artificial language learning. In J. Sprouse (Ed.), *Oxford handbook of experimental syntax* (pp. 271–300). Oxford University Press.

Damerow, P. (1996). Number as a second-order concept. *Science in Context, 9*(2), 139-149.

Davidson, K., Eng, K., & Barner, D. (2012). Does learning to count involve a semantic induction? *Cognition*, *123*(1), 162–173. https://doi.org/10.1016/j.cognition.2011.12.013

Döhler, C. (2018). *A Grammar of Komnzo* (Vol. 22). Language Science Press.

Fisher, C., Hall, D. G., Rakowitz, S., & Gleitman, L. (1994). When it is better to receive than to give: Syntactic and conceptual constraints on vocabulary growth. *Lingua*, *92*, 333–375.

Frank, M. C., Everett, D. L., Fedorenko, E., & Gibson, E. (2008). Number as a cognitive technology: Evidence from Piraha˜ language and cognition. *Cognition, 108*(3), 819–824.

Frye, D., Braisby, N., Lowe, J., Maroudas, C., & Nicholls, J. (1989). Young children's understanding of counting and cardinality. *Child Development*, 1158–1171.

Fuson, K. C. (2012). *Children's counting and concepts of number*. Springer Science & Business Media.

Fuson, K. C., Richards, J., & Briars, D. J. (1982). The Acquisition and Elaboration of the Number Word Sequence. In C. J. Brainerd (Ed.), *Children's Logical and Mathematical Cognition* (pp. 33–92). Springer New York. https://doi.org/10.1007/978-1-4613-9466-2_2

Gathercole, S. E. (1998). The development of memory. *The Journal of Child Psychology and Psychiatry and Allied Disciplines, 39*(1), 3-27.

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.

Gibson, D. J., Gunderson, E. A., & Levine, S. C. (2020). Causal Effects of Parent Number Talk on Preschoolers' Number Knowledge. *Child Development*, *91*(6). https://doi.org/10.1111/cdev.13423

Gleitman, L. (1990). The Structural Sources of Verb Meanings. *Language Acquisition*, *1*(1), 3–55. https://doi.org/10.1207/s15327817la0101_2

Gordon, P. (2004). Numerical cognition without words: Evidence from Amazonia. *Science, 306*(5695), 496–499.

Gould, P. (2017). Mapping the acquisition of the number word sequence in the first year of school. *Mathematics Education Research Journal*, *29*(1), 93–112.

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, 19*, 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. (2023). From partners to populations: A hierarchical Bayesian account of coordination and convention. *Psychological Review, 130*(4), 977.

Huang, Y. T., Spelke, E., & Snedeker, J. (2010). When Is *Four* Far More Than *Three* ?: Children's Generalization of Newly Acquired Number Words. *Psychological Science*, *21*(4), 600–606. https://doi.org/10.1177/0956797610363552

Hulme, C., Maughan, S., & Brown, G. D. (1991). Memory for familiar and unfamiliar words: Evidence for a long-term memory contribution to short-term memory span. *Journal of Memory and Language, 30*(6), 685-701.

Hurford, J. R. (1987). *Language and Number: The emergence of a cognitive system*. Oxford: Basil Blackwell.

Jara-Ettinger, J., Piantadosi, S., Spelke, E. S., Levy, R., & Gibson, E. (2017). Mastery of the logic of natural numbers is not the result of mastery of counting: Evidence from late counters. *Developmental Science*, *20*(6), e12459. https://doi.org/10.1111/desc.12459

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

Koponen, T., Aunola, K., & Nurmi, J.-E. (2019). Verbal counting skill predicts later math performance and difficulties in middle school. *Contemporary Educational Psychology*, *59*, 101803.

Krajcsi, A., Lengyel, G., & Kojouharova, P. (2016). The source of the symbolic numerical distance and size effects. *Frontiers in Psychology, 7*, 1795.

Krajcsi, A. (2017). Numerical distance and size effects dissociate in Indo-Arabic number comparison. *Psychonomic Bulletin & Review, 24*, 927-934.

Krajcsi, A., Lengyel, G., & Kojouharova, P. (2018). Symbolic number comparison is not processed by the analog number system: Different symbolic and non-symbolic numerical distance and size effects. *Frontiers in Psychology, 9*, 124.

Le Corre, M., & Carey, S. (2007). One, two, three, four, nothing more: An investigation of the conceptual sources of the verbal counting principles. *Cognition*, *105*(2), 395–438.

Le Corre, M. (2014). Children acquire the later-greater principle after the cardinal principle. *British Journal of Developmental Psychology, 32*(2), 163-177.

Le Corre, M., Li, P., Huang, B. H., Jia, G., & Carey, S. (2016). Numerical morphology supports early number word learning: Evidence from a comparison of young Mandarin and English learners. *Cognitive Psychology*, *88*, 162–186.

LeFevre, J.-A., Skwarchuk, S.-L., Smith-Chant, B. L., Fast, L., Kamawar, D., & Bisanz, J. (2009). Home numeracy experiences and children's math performance in the early school years. *Canadian Journal of Behavioural Science/Revue Canadienne Des Sciences Du Comportement*, *41*(2), 55.

Marchand, E., Lovelett, J. T., Kendro, K., & Barner, D. (2022). Assessing the knower-level framework: How reliable is the Give-a-Number task? *Cognition*, *222*, 104998.

Marques, W. (2022). Multinumeramento em tela: esmiuçando o constructo de múltiplas representações. *Boletim GEPEM, (80)*, 119-142.

Martin, R. B., Cirino, P. T., Sharp, C., & Barnes, M. (2014). Number and counting skills in kindergarten as predictors of grade 1 mathematical skills. *Learning and Individual Differences*, *34*, 12–23.

Marušič, F., Saksida, A., Sullivan, J., Skordos, D., Wang, Y., & Barner, D. (2021). Do children derive exact meanings pragmatically? Evidence from a dual morphology language. *Cognition*, *207*, 104527.

Marušič, F., Žaucer, R., Plesničar, V., Razboršek, T., Sullivan, J., & Barner, D. (2016). Does grammatical structure accelerate number word learning? Evidence from learners of dual and non-dual dialects of Slovenian. *PloS One*, *11*(8), e0159208.

Meyer, C., Barbiers, S., & Weerman, F. (2020). Many systems, one strategy: Acquiring ordinals in Dutch and English. *Glossa: A Journal of General Linguistics*, *5*(1). https://www.glossa-journal.org/article/id/5348/

Mix, K. S., Sandhofer, C. M., Moore, J. A., & Russell, C. (2012). Acquisition of the cardinal word principle: The role of input. *Early Childhood Research Quarterly, 27*(2), 274-283.

Negen, J., & Sarnecka, B. W. (2012). Number-Concept Acquisition and General Vocabulary Development. *Child Development*, *83*(6), 2019–2027. https://doi.org/10.1111/j.1467-

8624.2012.01815.x

Nikoloska, A. (2009). Development of the cardinality principle in Macedonian preschool children. *Psihologija*, *42*(4), 459–475.

Overmann, K. A. (2018). Constructing a concept of number. *Journal of Numerical Cognition, 4*(2), 464-493.

Overmann, K. A. (2023). The materiality of numbers: Emergence and elaboration from prehistory to present. Cambridge University Press.

Piantadosi, S. T., Jara-Ettinger, J., & Gibson, E. (2014). Children's learning of number words in an indigenous farming-foraging group. *Developmental Science*, *17*(4), 553–563. https://doi.org/10.1111/desc.12078

Pica, P., Lemer, C., Izard, V., & Dehaene, S. (2004). Exact and approximate arithmetic in an Amazonian indigene group. *Science, 306*(5695), 499–503.

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

Pitt, B., Gibson, E., & Piantadosi, S. T. (2022). Exact number concepts are limited to the verbal count range. *Psychological Science, 33*(3), 371-381.

Sarnecka, B. W., & Carey, S. (2008). How counting represents number: What children must learn and when they learn it. *Cognition*, *108*(3), 662–674. https://doi.org/10.1016/j.cognition.2008.05.007

Sarnecka, B. W., Kamenskaya, V. G., Yamana, Y., Ogura, T., & Yudovina, Y. B. (2007). From grammatical number to exact numbers: Early meanings of 'one','two', and 'three' in English, Russian, and Japanese. *Cognitive Psychology*, *55*(2), 136–168.

Schlimm, D., & Neth, H. (2008). Modeling ancient and modern arithmetic practices: Addition and multiplication with Arabic and Roman numerals. In *30th Annual Meeting of the Cognitive Science Societ*y.

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.

Schneider, R. M., Pankonin, A., Schachner, A., & Barner, D. (2021a). Starting small: Exploring the origins of successor function knowledge. *Developmental Science, 24*(4), e13091.

Schneider, R. M., Sullivan, J., Guo, K., & Barner, D. (2021b). What Counts? Sources of Knowledge in Children's Acquisition of the Successor Function. *Child Development*, *92*(4). https://doi.org/10.1111/cdev.13524

Siegler, R. S., & Robinson, M. (1982). The development of numerical understandings. In *Advances in child development and behavior* (Vol. 16, pp. 241–312). Elsevier. https://www.sciencedirect.com/science/article/pii/S0065240708600725

Spaepen, E., Gunderson, E. A., Gibson, D., Goldin-Meadow, S., & Levine, S. C. (2018).

Meaning before order: Cardinal principle knowledge predicts improvement in understanding the successor principle and exact ordering. *Cognition*, *180*, 59–81.

Turan, E., & De Smedt, B. (2022). Mathematical language and mathematical abilities in preschool: A systematic literature review. *Educational Research Review*, *36*, 100457.

Sella, F., & Lucangeli, D. (2020). The knowledge of the preceding number reveals a mature understanding of the number sequence. *Cognition, 194*, 104104.

Wagner, K., Chu, J., & Barner, D. (2019). Do children's number words begin noisy? *Developmental Science*, *22*(1), e12752. https://doi.org/10.1111/desc.12752

Wagner, K., Kimura, K., Cheung, P., & Barner, D. (2015). Why is number word learning hard? Evidence from bilingual learners. *Cognitive Psychology*, *83*, 1–21.

Wege, T. E., Bourque, T., Merkley, R., & Cheung, P. (2022). *Thirty years of knower-levels: A systematic review of the Give-N task*. https://psyarxiv.com/fh95s/download?format=pdf

Willits, Jones, & Landy. (2016). Learning that numbers are the same, while learning that they are different. In *38th Annual Meeting of the Cognitive Science Societ*y.

Wright, R. J. (1994). A study of the numerical development of 5-year-olds and 6-year-olds. *Educational Studies in Mathematics*, *26*(1), 25–44.

Wynn, K. (1990). Children's understanding of counting. *Cognition*, *36*(2), 155–193.

Wynn, K. (1992). Children's acquisition of the number words and the counting system. *Cognitive Psychology*, *24*(2), 220–251.

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

Supplementary Materials

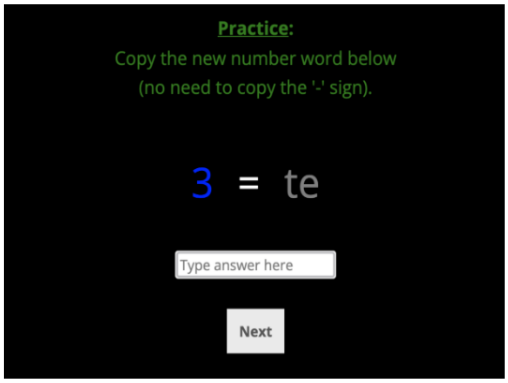
The Instructions page for Experiment 1. Instructions for Experiment 2 were similar, but also mentioned that participants would then use the numbers in 2 new tasks, with no feedback.

Instructions

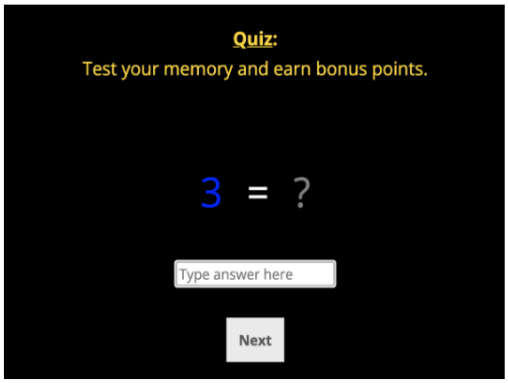
Number System

In this experiment you will be learning a new number system. Your task is to practise and name numbers as quickly and accurately as possible. You will complete 65 trials in total.

Sometimes before the trial, you will be shown a new number word and asked to type it out for practice. For example, here is a new word equal to '3'. You would write 'te' in the box:



In most trials, you will just be presented with a number and asked to provide a translation for it in the new number system:



Keep in mind that if there is a hyphen (-) in the number word, you don't have to type it in your answer. So for example, if the number's name is 'te-ko', you could just write 'teko'.

Finally, please DO NOT refresh the page at any point once you start the game. If you refresh, you will not be able to get back to the session you were in, you will lose your progress, and you may not be able to submit the task.
Thank you!

Please fill out this short quiz below before getting started. Good luck and have fun!

Comprehension Quiz

If you refresh the page at any point, you will lose all progress.
- ○ False
- ○ True

If the number word has hyphens, you will have to type the hyphens in your response.
- ○ False
- ○ True

Submit

Chapter 3

Can preschoolers learn the syntax of number? Using rules to combine familiar and novel number words

Sebastian Holt & David Barner

Department of Psychology, University of California, San Diego

Abstract

Although young children can combine words to express compositional meanings like "red house" or "two cats", they do not represent number words compositionally (e.g., "sixteen" as 6+10) until much later in development. One reason for this might be that smaller number words in most languages (including English) do not contain compositional rules, limiting children's exposure to the syntactic structure underlying larger number words. However, historically, many languages have featured number systems with smaller bases or anchors, thus relying more on rules to represent smaller numbers. Might children acquire rules for combining numbers earlier in development, if given evidence of such rules in their input? We explored this in two experiments asking how children interpret expressions composed of small number words. In Experiment 1, we presented children with a system anchored at 2, and found that many children could compose small number words using a conjunctive rule - e.g., "Give me two and one apples", and some older children could also comprehend novel multiplicative expressions - e.g., "Give me two *twos* of bananas". In Experiment 2, we found that children applied additive rules even to a system of novel number words ("Monkey numbers") anchored at 3.

Keywords: number; counting; syntax; compositionality; base

To express large, exact numbers, most languages combine small numbers through a system of generative rules (Comrie, 2013; Hurford, 1987). For example, the Mandarin word for "twenty-six", *èr-shí-liù*, is composed directly of words meaning 'two', 'ten', and 'six', which are added or multiplied together based on their syntactic position within the expression. Knowledge of these syntactic rules may allow children to produce and conceptualize indefinitely large numbers (Cheung et al., 2017; Chu et al., 2020; Falk, 2010; Guerrero et al., 2020; Sullivan et al., 2023), and to reason arithmetically about numbers they've never encountered before (Buijsman, 2020; Cheung et al., 2016; Secada, Fuson, & Hall, 1983; Sullivan et al., under review). Interestingly, although children acquire various morpho-syntactic connectives in their first 2-3 years of life (Bloom et al., 1980; Lust & Mervis, 1979; Fenson et al., 1994), they don't understand the syntactic rules that combine number words until at least age 5 or 6 - several years after beginning to use them in speech (e.g., Schneider et al., 2020; 2021). In the present study, we considered the possibility that children may be capable of learning such rules early in development, but fail to learn them because the numbers they are most likely to hear and use – e.g., words for 1-20 – lack clear compositional structure. To explore this possibility, we conducted two experiments in which we exposed children to number systems with "anchors"[18] of 2 or 3, which required compositional rules to express even relatively small number words.

Children's earliest knowledge of number words comes from two largely separate sources. One source is associations between individual number words and cardinalities. Beginning at around age 2, English-speaking children in the US begin to learn that the words *one*, *two*, and *three* can refer to small sets of 1, 2, and 3 objects, respectively (and are sometimes called "subset

[18] In this special issue we use the term "anchor" to refer to a compositional structure that acts like a base but does not have special symbols for powers of a number.

knowers", since they know the meanings of only a few number words; Wynn, 1990, 1992). Thus, children begin with a finite, non-compositional, number system, and can only label and construct sets up to a maximum of 3-4 objects. While the age at which children make these item-based associations varies across children, languages, and cultures, multiple studies find that they are nevertheless acquired in the same, ascending, order (Almoammer et al., 2013; Barner et al., 2009; Jara-Ettinger et al., 2017; Le Corre & Carey, 2007; Le Corre, Li, Huang, Jia & Carey, 2016; Marušič et al., 2016; Piantadosi, Jara-Ettinger & Gibson, 2014; Sarnecka & Carey, 2008; Sarnecka et al., 2007; Sarnecka & Lee, 2009; Schneider et al., 2020; Wynn, 1990, Wynn, 1992; Wagner et al., 2015). A second source of number knowledge is the counting procedure. When asked to give a set of three objects, subset knowers seldom spontaneously count and are unable to accurately give large numbers (e.g., Wynn, 1990; Le Corre et al., 2006). For most subset knowers, counting is a meaningless, rote procedure, the logic of which they do not understand or associate reliably with numerosity. However, by around age 3½ or 4, children appear to have an insight about counting, and realize that the final number uttered while counting a set is also the cardinal label of that set - i.e., the cardinal principle (Wynn, 1990; 1992). With this discovery, they become what is termed "cardinal principle knowers" (or CP knowers).

Recent studies have found that after learning how to accurately count large sets, children continue to learn basic properties of counting for several years (Le Corre, 2014; Davidson, Eng, & Barner, 2012; Schneider et al., 2020; Spaepen et al., 2018). One late emerging skill is children's mastery of the rules that govern the combination of number words. In base-10 number systems like English, most numbers from 1-99 are composed by concatenating the unit labels from *one* through *nine* with decade labels like *twenty* and *thirty* - e.g., generating *twenty-three* from *twenty* and *three.* Meanwhile, some numbers are irregular and must be memorized, such as the English

teens, and the decade labels themselves. In other languages, like Mandarin, there are no irregularities in the teens, and the decade labels can be composed through transparently multiplicative rules - e.g., *twenty* is expressed as *èr shí* - literally "two ten" - and *twenty three* is *èr shí sān*, or "two ten three". Previous studies suggest that across languages, children don't learn these rules until between the ages of 4 and 6 - well after they become CP-knowers - and that until this time children have a finite, memorized, and unstructured count list. One key piece of evidence for this comes from children's errors when asked to count as high as they can. If children were to rote-memorize numbers in their count list, then we would expect their errors to occur randomly, reflecting the frequency of exposure they have to counting to particular numbers - e.g., stopping at 12, or 24, or 37. For very young children - e.g., most subset knowers, and some young CP-knowers - this is what we find. However, the most common errors among older English-speaking CP knowers are not random, but tend to be numbers ending in *nine* - e.g. 29, 39, 49, etc. (Fuson et al., 1982; Gould, 2017; Siegler and Robinson, 1982; Schneider et al., 2020; Wright, 1994). Such errors suggest that children count by composing decade labels with the units 1-9, and fail when they reach decades like *forty* and *fifty*, which depend on rote memory. In fact, even in languages where decade labels are more transparently composed using rules (e.g., Mandarin), children also seem to master decade + unit compositions before they show sensitivity to the composition of decade labels (Guerrero et al., 2020; Miller & Stigler, 1987; Miura & Okamoto, 1989; Miura et al., 1988, 1993; Schneider et al., 2020).

One possible explanation of these developmental facts is that children acquire rules for combining number words late because this ability depends on a prior mastery of counting, and the principles that underlie it (e.g., see Spaepen et al., 2018, for a related proposal). For example, in order to notice that *twenty-three* is composed from *twenty* and *three* children might first need to

learn to count to this number, assign it a cardinal value, recognize that *twenty-three* is one more than *twenty-two*, etc. An alternative possibility, however, is that children might learn these rules earlier if they received greater exposure to them in their input. In base-10 languages like English, the compositional rules of number words are not obvious in the first 30 number words, which many children do not begin to use prior to the age of 4 or 5, and hear infrequently in caregiver speech (Willits, Jones, & Landy, 2016). For example, in the English number system, the first rule-like repetitions of the decade+unit rule (e.g., *twenty-five; thirty-five,* etc.) only occur when counting into the 30s. Whereas most of what we know about number word learning comes from studying base-10 systems, the cultural history of number suggests that this is only one of many possible systems, and that many alternative systems are attested on the historical record, featuring bases or anchors ranging from 2 to 20 (Comrie, 2005; Hammarström, 2010; WALS). For example, in the number system historically attested in the Torres Strait (Haddon, 1890), there were only words for one and two – *urapun* and *okosa* – and rules for combining them additively to create larger numbers: *okosa-urapun* (i.e., 3), *okosa-okosa* (i.e., 4), *okosa-okosa-urapun* (i.e., 5), etc. (for other examples, see Donahue, 2008; Epps, 2006; Epps et al., 2012). This system was not used for counting, and was seldom used to express quantities larger than 6, for which a separate, non-compositional, body-based tally system was used. This raises the question of whether children exposed to such a system might acquire compositional rules earlier, and without requiring mastery of counting procedures.

Previous studies have not directly tested whether children are capable of learning rules for combining numbers prior to learning to count. However, there is evidence that young children might have the ingredients required for this capacity. First, previous studies suggest that children as young as 2 can use the word "and" to conjoin noun or verb phrases within a sentence, e.g., "you

can carry that and I can carry this" (Bloom et al., 1980; Morris, 2008). Second, as noted by Spelke (2017), if children's representation of "and" were sufficiently abstract to generalize to novel syntactic contexts (i.e., number words), children might use this operator to construct large number word meanings from small ones - e.g., coming to understand the meaning of "five" through complex noun phrases like *three dogs and two cats*. According to Spelke:

> "...children learn natural language expressions that contain terms for individuals of different kinds (the dog and the cat) or individuals with different identities (Evelyn and Simon; my cup and your cup; this dog, that dog, and the other one). Using conjunction, these expressions refer to sets of two or three distinct individuals. They also support learning of new nouns at different taxonomic levels (e.g., Look at those birds: there's a duck, a goose and a swan). Finally, these expressions support learning of number word meanings (e.g., John has three pets: a dog, a cat, and a hamster). Thus, children can learn part of the meanings of the first number words as they occur in simple noun phrases referring to objects (two dogs), and eventually to entities of other kinds (two oinks, three jumps). Children discover that such noun phrases, like noun phrases with two or three conjuncts (the big dog and the little one), refer to two or three distinct individuals." (Spelke, 2017, p. 157)

Related to this idea, the cultural history of number suggests that when compositional rules emerge in number systems, they often encode operations explicitly via words like "and", as shown in (1-5), among many examples. In some cases, these expressions endure, while in others they become simplified over time and lose these explicit markers (e.g., English, which used to compose *one-and-twenty* analogously to Dutch). This gradual loss of explicit morphology might make the

meaning of compositional expressions opaque to those who are not familiar with context that gave rise to them, but also make them easier to use (Pickering & Garrod, 2004; Cooperrider & Gentner, 2019; Hawkins et al., 2023). The existence of explicit conjunction in emerging number systems suggests two conclusions. First, humans may naturally conceive of large numbers as syntactic compositions of smaller ones. And second, the use of explicit morphemes like conjunction may be an important stepping stone for language users - and perhaps young children - to understand this compositional structure.

(1) *een en twintig* (Dutch, lit. "one and twenty")

(2) *irupatti onru* (Tamil, lit. "twenty.and one")

(3) *amashumi amabili nanye* (Zulu, lit. "ten two and.one")

(4) *un ar hugain* (Welsh, lit. "one and twenty")

(5) *vingt et un* (French, lit. "twenty and one")

A challenge to testing this idea is that most currently attested number systems feature a base-10 structure. Consequently, it is difficult to know from existing studies whether children might learn compositional rules earlier if they were exposed to a system with a smaller base. That said, a small number of past studies have tested children using novel, artificial, number systems. For example, in one study, Cheung et al. (2016) asked whether 4- to 6-year-old children could learn a novel multiplier word (*gobi*, meaning "three") and generalize its use at test when combined with a numeral - e.g., "Who has two gobi books?". They found that both English-speaking and Cantonese-speaking children successfully comprehended these multiplicative constructions. A limitation of this study, however, is that participants were relatively old, making it unclear whether

younger children, who lack prior mastery of counting, could also learn rules for combining numbers. Thus, the study did not address whether acquisition of rules for combining numerical expressions depends on earlier stages of numerical development, or if instead rules can be learned earlier.

To address this question, we conducted two experiments to test young children's ability to comprehend novel combinations of small number words. In Experiment 1, we presented children with rules that combined the familiar English numbers *one* and *two*, either additively or multiplicatively. Additive rules either featured explicit conjunctive morphology - e.g., "Give me two and one bananas" - or implicit conjunction - e.g., "Give me two-one bananas". Multiplicative requests asked children to, e.g., "Give me two twos of bananas", or like Cheung et al. (2016) used a measure expression "lunch" which was associated with two food items - e.g., "Give me two lunches of bananas." To explore the effects of age and counting experience, we tested children from 2 to 6 years of age, including subset knowers, who had not yet learned to count, and CP knowers, who had. In Experiment 2, we extended this logic to an entirely novel number system with an anchor of 3, by teaching children words for 1 (*gi*) and 3 (*zo*) and asking them to interpret combinations of these words in additive constructs, as in Experiment 1.

Experiment 1

We presented children with a novel number system, in which small familiar English number words were composed to express larger numbers using either an additive or multiplicative rule. Using this method, we asked whether children could spontaneously comprehend these compositional rules, and also whether their ability to do so was related to their mastery of counting (by testing both subset knowers, who could not accurately count large sets, and CP knowers, who

could). We also asked whether the presence of an explicit marker, "and", would help children recognize the intended compositional meaning of each expression.

Methods

Pre-registrations for Experiments 1 and 2 can be found at: https://osf.io/w6jtq/. Data and analysis code can be found at: https://github.com/SebastianHolt/number_rules

Participants

We recruited 162 English-speaking children at preschools and museums in San Diego, California (N=89), and Comox, British Columbia (N=73). We excluded 35 children based on pre-registered criteria (28 for not finishing the experiment; 4 due to Give-a-Number classification errors; 1 repeat participant; 1 due to missing age data), and 8 further children who were accidentally recruited in excess of the preregistered limit, leaving a final sample of 120 (age range 2;0 to 6;11; M = 4;1; SD = 0;11).

Design

In the two Conjunction tasks, we manipulated the presence or absence of the word "and" within composed number words (e.g., "two and two"). In the Multiplier task, we manipulated the use of either a number, "two", or a noun, "lunch", in a measure phrase denoting two objects. Subset and CP knowers were assigned to each condition.

Procedure

Give-a-Number. Children first received a titrated version of the Give-a-Number task (Marchand et al., 2022). Starting with the number *five*, children were asked to give the experimenter *n* fish. If they responded correctly for *n* they were asked for *n*+1. When they responded incorrectly, they were asked for *n*-1 (or *n* again if *n* was *one*). Children's knower level was defined as the highest number for which they correctly provided *n* on at least 2 out of 3 trials, while not providing *n* more than once for other requests. Children who knew *five* were credited as CP knowers. Using this method we identified 3 non-knowers, 6 1-knowers, 20 2-knowers, 22 3-knowers, 9 4-knowers, and 60 CP-knowers.

Highest Count. This probed children's familiarity with numbers up to 20. Children were asked to count as high as they could and were stopped after reaching 20 (or higher when time permitted). Because highest count was capped at 20 for many children, it was not included as a predictor in our analyses.

Small Set Conjunction Task. In this task, children made juice recipes for a puppet, Elephant, who left instructions in "Elephant language". All recipes required taking fruits from two containers and mixing them in a cup. The experimenter first demonstrated one example recipe, which they described as "this many *and* this many apples" (Explicit Conjunction condition) or "this many, this many apples" (Implicit Conjunction condition). While saying each term of the compound expression, the experimenter drew a set of apples first from one container, and then the other, allowing children to associate each term with a distinct step in the recipe. Children were then asked for sets of fruits using expressions of three different forms (1+1, 2+2, and 2+1), for each of three different fruits (apples, oranges, and strawberries), in that order, with no feedback - e.g., "Could you make *two* and *one* apples?". In both conditions, these requests were expressed using the English number words *one* and *two*. In the Implicit Conjunction condition, the additive

operation was only implied, so requests were expressed as *one-one apples*, *two-two apples*, or *two-one apples*. In the Explicit Conjunction condition, the additive operation was expressed with the word "and", i.e., *one-and-one apples*, *two-and-two apples*, or *two-and-one apples*. We recorded how often children returned the sum of the two composed number words (e.g., giving 3 for 2+1 requests).

Large Set Conjunction Task. Next, CP knowers (but not subset knowers) completed a similar set of trials for larger numbers. The task tested whether CP knowers could reason about the conjunction of large sets constructed via counting. Only CP knowers were included, since only they were able to construct sets as large as 10. Children were presented with two open containers of 10 apples each, and a large cup. Trials began with a brief comprehension check, where the experimenter showed one recipe: 10 apples. Children were then asked to repeat the experimenter's action. Next they were asked for sets of 10+*n* apples where *n* included 1-9 (i.e., sets of 11, 12, …, 19). Requests mirrored those in the Small Set Conjunction task, such that 10+1 apples was *ten-one apples* in the Implicit Conjunction condition, and *ten-and-one apples* in the Explicit condition.

Multiplier Task. Analogous to Cheung et al. (2016), this task asked whether children spontaneously interpret digit-multiplier expressions multiplicatively. To do so, we told children that Elephant had drunk enough juice, and now wanted food. They were shown what "a lunch of bananas" looks like (Nominal Condition) or "a two of bananas" (Numeral condition), which in both cases was two bananas in a small cup. Two negative examples were also provided: when a banana was removed, children were told that it was *not* a lunch/two of bananas and Elephant would remain hungry. When an extra banana was added, children were told that this was *not* a lunch/two of bananas and Elephant would have a stomach ache because there were too many. Children saw

four lunches of bananas created, and were then prompted to give "one of those *lunches"* (Nominal Condition) or "one of those *twos"* (Numeral Condition). Once they had, they were then asked to give "two of the lunches" or "two twos", and given feedback if necessary.[19] Next, in the test phase, children were presented with a set of bananas and were asked to create "one lunch of bananas" (Nominal Condition) or "one two of bananas" (Numeral Condition). They were then asked to create "two lunches of bananas" (Nominal) or "two twos of bananas" (Numeral). The experimenter then repeated this procedure separately for strawberries, including the training phase. Finally, the cups were removed and children created sets directly from fruits, and all trials were repeated.

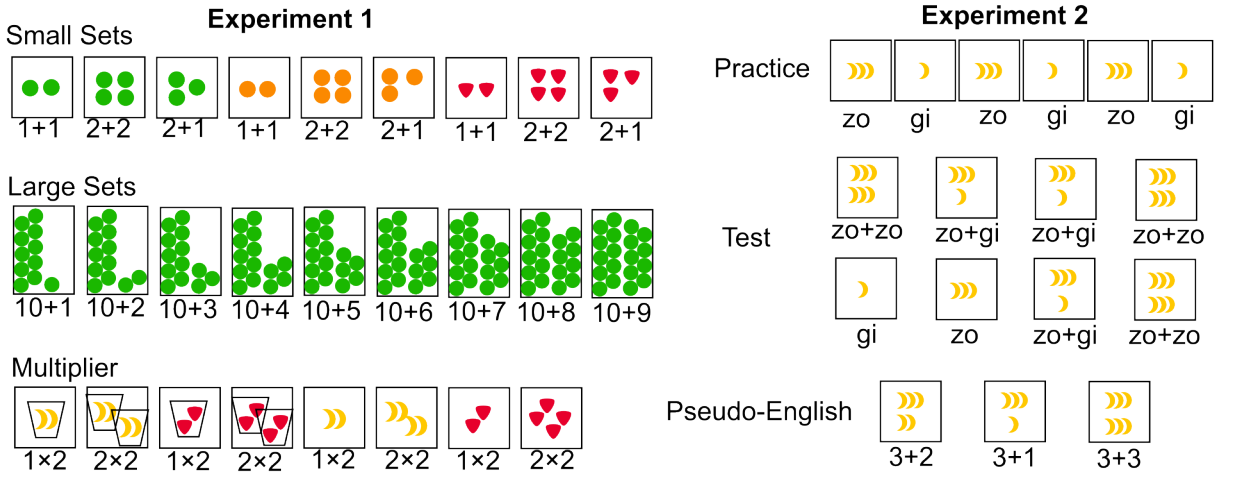


Figure 3.1. Trial structure of each task in Experiments 1 and 2. Depicted is the expected, "target" set of each trial. N.B. only CP knowers completed the Large Sets trials. Other than the second term of each Multiplier task request, Arabic numerals may be read as English number words, and the + operator as either "and" or silent.

Results

[19] We originally pre-registered that we would record the number of feedback iterations required, to be used as a predictor in models. However, enumerating feedback turned out to be impossible, since often a single instance of feedback required multiple explanations that often were not separable. We therefore stopped recording this measure and omitted it from models.

Small Set Conjunction Task. We first examined children's performance in the Small Set Conjunction task. To do so, we constructed a mixed-effects linear regression predicting accuracy on each trial, with condition (Implicit Conjunction vs. Explicit Conjunction), target quantity (2, 3, or 4), age (in months), and CP knower status (True or False), as predictors, as well as a random intercept for each child. Data were analyzed using R (Version 4.3.3, R Core Team, 2024) and using the lme4 package (Bates et al., 2015). Children's responses were significantly more accurate in the Explicit Conjunction condition (54.4 ± 4.2%), than in the Implicit Conjunction condition (36.9 ± 4.1%; b=1.4, z=3.682, p<0.001; see Figure 3.2). Children were also most accurate for the smallest target quantity (i.e., 2), and less accurate as the target quantity increased to 3 or 4 (b=-0.393, z=-3.742, p<0.001). CP knowers provided more accurate responses than subset knowers in both the Explicit condition (subset knowers: 35.9% ± 5.7%; CP knowers: 73% ± 5.3%) and the Implicit condition (subset knowers: 18.1% ± 4.6%; CP knowers: 55.6% ± 5.9%). Accordingly, CP knowledge was a significant predictor of accuracy (b=1.142, z=2.322, p=0.02). Finally, we found a significant effect of age (b=0.093, z=3.937, p<0.001).

Separate exploratory analyses for subset and CP knowers found that there was an effect of age for CP knowers (b=0.126, z=3.269, p=0.001), as well as for subset knowers (b=0.06, z=2.016, p=0.044). Also, both subset knowers (b=1.247, z=3.088, p=0.002) and CP knowers (b=1.588, z=2.27, p=0.023) showed higher accuracy for requests with "and" than for those without. This provides strong evidence that some subset knowers deployed rules for composing number words, and that they did so more in the Explicit Condition. We next asked whether subset knowers of different knower levels (0, 1, 2, 3, or 4) benefitted differently from the presence of "and". We repeated the above model predicting subset knowers' accuracy from condition, target quantity, and age, and added knower-level (0-4) as a predictor. Knower level predicted accuracy (b=0.442,

z=2.168, p=0.03), while age did not (b=0.049, z=1.705, p=0.088). Finally, we asked whether subset knowers were more likely to provide additive responses if the target quantity was within their knower-level (e.g., 3-knowers responding to "two-and-one") or outside of their knower level (e.g., 3-knowers responding to "two-and-two"). This effect was not significant (b=0.185, z=0.524, p=0.601), suggesting that the effect of knower level was not exclusively due to already knowing the word for the sum of the component numbers.

We next explored alternative strategies children used in this task. In addition to correct responses, which we call the "Additive" strategy, we identified four other common strategies (see Table 3.1). Here we report children who were consistent with one strategy on at least 7/9 trials (67 of 120 children). In an Indexing strategy, each number word could be interpreted as the label for one object in the set - e.g., such that "*two* and *two*" meant "*an object* and *an object*". Because all requests contained two number words, this would predict a response of 2 on every trial.[20] In a First Word or Last Word strategy, children might only attend to the first or last number word in an expression - e.g., giving 1 when asked for "two and one" - ignoring the other word. Finally, a Give Many strategy involved simply giving a large number of objects on every trial, which we operationalized as any number greater than 6.

Around one third of subset knowers used a consistent strategy in both the Explicit (10/30) and Implicit conditions (11/30). Half of consistent subset knowers in the Explicit condition used an Additive strategy (5/30), while none did in the Implicit condition, instead responding most often with a First Word (4/30) or Last Word strategy (3/30). They also used the Indexing strategy in both the Explicit (3/30) and Implicit (2/30) conditions. Finally, two subset knowers used the Give Many strategy in both the Explicit (2/30) and Implicit (2/30) conditions.

[20] In contrast, no child responded to all 9 trials with a quantity other than 2.

CP knowers were more consistent in both the Explicit Conjunction condition (23/30) and Implicit Conjunction condition (23/30). Most used the Additive strategy in both the Explicit (19/30) and Implicit (13/30) conditions. Several used the Indexing strategy in the Explicit (3/30) and Implicit (5/30) conditions. They also used a First Word strategy in the Implicit condition (3/30), but never the Explicit condition, and none used a Last Word strategy in either condition. Some also used the Give Many strategy in the Explicit (1/30) and Implicit (2/30) conditions

Table 3.1. The number of subset and CP knowers in each condition who were consistent with alternative strategies on at least 7/9 trials in each of the Small and Large Sets tasks.

Small Sets Task		**Additive**	**Indexing**	**First Word**	**Last Word**	**Give Many**
Subset	Explicit	5	3	0	0	2
	Implicit	0	2	4	3	2
CP	Explicit	19	3	0	0	1
	Implicit	13	5	3	0	2
All		37	13	7	3	7
Large Sets Task						
CP	Explicit	12	0	4	0	1
	Implicit	11	1	4	0	5
All		23	1	8	0	6

Large Set Conjunction Task. CP knowers gave accurate responses for conjunctions of large numbers 51.1% (± 6%) of the time in the Explicit Conjunction condition, and 39.6% (± 5.8%) of the time in the Implicit condition (see Figure 3.2). To explore these results, we constructed a model similar to the one used to predict responses on small number trials (but without CP knowledge as a predictor, since subset knowers did not complete this task). This model did not show a main effect of condition (b=2.07, z=1.692, p=0.09) suggesting that CP knowers were similarly likely to interpret compositions additively whether conjunction was marked explicitly or implicitly. Also, there was an effect of set size, such that children gave less accurate responses for larger numbers (b=-0.167, z=-2.683, p=0.007), and an effect of age, with older CP knowers providing more accurate responses than younger CP knowers (b=0.199, z=3.005, p=0.003).

Using the same criteria as in the Small Sets Conjunction task, we identified 12 children in the Explicit condition and 11 children in the Implicit condition as consistent responders with the Additive strategy. Further 4 were consistent with a First Word strategy in each condition, and 1 child in the Implicit condition was consistent with the Indexing strategy. Responses of all 20 apples were coded as Give Many, which characterized 1 child in the Explicit condition and 5 in the Implicit condition.

Multiplier Task. In this task, we found that CP knowers provided correct responses on 56.7% ± 4.4% of trials, while subset knowers did so 32.5% ± 4.2% of the time (see Figure 3.2). CP knowledge predicted higher accuracy (b=1.123, z=2.886, p=0.004), whereas age did not (b=0.027, z=1.526, p=0.127). The smaller target quantity 2 was easier for children than 4 (b=-1.037, z=-11.162, p<0.001), and requests were also easier when the word used to express a set of two fruits was “two” (49.4% ± 4.5%), rather than “lunch” (39.8% ± 4.4%), (b=0.664, z=2.243, p=0.025). This difference between conditions may be because some CP knowers interpreted the

expression “two twos” additively, recycling a successful strategy from the Small Set Conjunction Task.

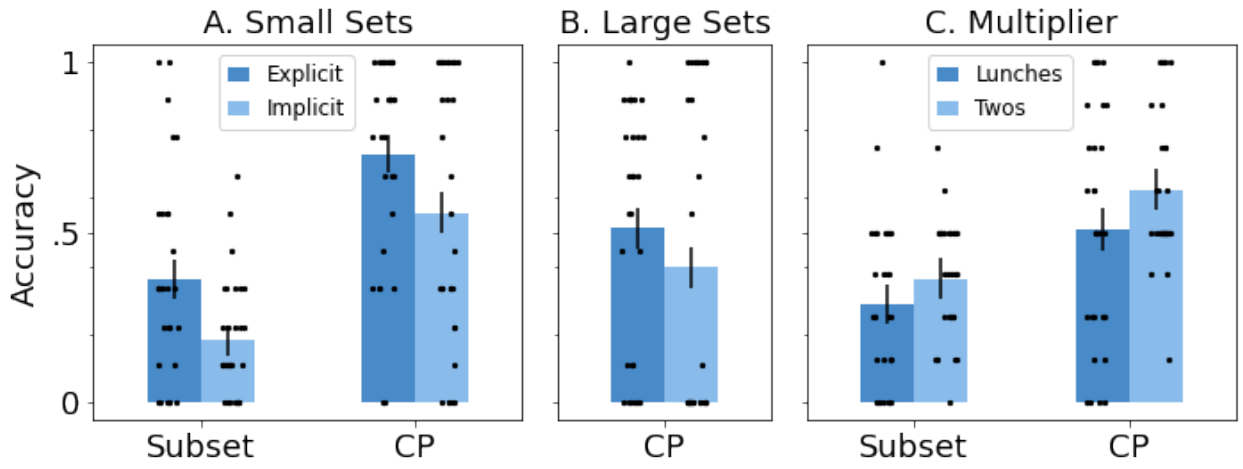


Figure 3.2. A. Subset and CP knowers’ accuracy in Small Sets task. B. Only CP knowers completed the Large Sets task. C. Multiplier task. Error bars are 95% confidence intervals. Dots represent individual children.

Discussion

Experiment 1 found three main results. First, subset knowers often gave additive interpretations when familiar numbers were conjoined with “and”. Second, CP knowers also interpreted such requests additively, but also did so (1) without the support of “and”, (2) for sets larger than 10, and (3) increasingly often with age. Finally, CP knowers often gave accurate replies for multiplicative expressions, and were more accurate than subset knowers on this task. These results provide a preliminary demonstration that children can compose existing number words using familiar combinatorial rules to construct larger quantities. They therefore suggest that even young children may have the capacity to understand the syntactic structure of numerical systems, and that the late emergence of such rules in observational datasets may reflect the lack of such rules in children’s input.

In Experiment 2, we sought to extend this result to a truly novel number system, in which children learned novel labels for small sets. Interestingly, previous work finds that when bilingual children learn more than one number system - e.g., Spanish and English - knowledge of small number word meanings does not transfer across languages (resulting in different "knower levels"), while understanding of the cardinal principle does (Wagner et al., 2015), despite significant differences in exposure to these different number systems. In this spirit, in Experiment 2 we compared CP knowers' ability to use conjunctive rules in both pseudo-English numbers as in Experiment 1, and "Monkey numbers" composed of novel number words. We asked whether children could deploy rules for novel numbers, and also whether, like CP knowledge, the ability to use combinatorial rules expresses itself equally across number languages, or if instead it differs for novel vs. familiar number words.

Experiment 2

In Experiment 2, we trained CP knowers on "Monkey numbers", including novel labels for 1 (*gi*) and 3 (*zo*). We then asked them to interpret expressions that conjoined these labels with or without "and". Finally, we tested them with equivalent compositions of English numbers, as in Experiment 1. Our main questions were (1) whether children could apply conjunctive rules to newly learned number words in a system with an anchor of 3, and (2) whether performance on English numbers would predict performance on novel numbers.

Methods

Participants

We recruited 90 English-speaking children at preschools and museums in San Diego, California, of whom 30 were excluded on the basis of pre-registered criteria (8 subset knowers; 18 did not complete the task; 4 collected after we had reached our planned sample size), leaving a sample of 30 per condition (M_age = 57.37 months; SD_age = 8.74 months).

Design

As in Experiment 1 we tested children in an Explicit Conjunction and an Implicit Conjunction condition. Also, there were two within subjects conditions: A novel number word condition, followed by an English number word condition.

Procedure

Give-a-Number. All children first received a titrated version of the Give-a-Number task, as in Experiment 1, to identify CP knowers for inclusion in the study.

Monkey Numbers. In the Training Phase, children were introduced to a monkey puppet and told that they were going to learn "monkey numbers". The experimenter showed children "*zo* bananas", setting out three plastic banana counters onto a plate. Children were asked how many bananas there are when you have *zo* bananas, and then prompted to recreate *zo* bananas themselves. The same demonstration was repeated for "*gi* banana", which was a set of one. Next, children were given practice with each number in the "fast game". In six trials, the experimenter said "*zo*" or "*gi*" and the child responded with a set of bananas. After each incorrect response, children were prompted to fix their mistakes. If they refused or did so incorrectly, the experimenter provided an example of the correct number. These trials always occurred in the same sequence: *zo, gi, zo, gi,*

zo, gi. Only children who responded correctly to each of the last two trials continued to the test trials.

In the Test Phase, children were told they would play a guessing game with the monkey numbers, in which the experimenter said a new monkey number that the child had not heard before, and the child was asked to make a set of bananas in response. Children were asked for *zozo bananas* (Implicit Conjunction condition) or *zo and zo bananas* (Explicit Conjunction condition), without feedback. The Training Phase demonstrations for *gi* and *zo* were then repeated (to prevent decay of memory for these words), and then children played the guessing game again for *zogi* or *zo and gi*. After another repetition of the Training Phase demonstration, they were asked for *zogi* and *zozo* again. Then, the experimenter introduced a giraffe puppet, who was also learning the monkey numbers. Children were asked to teach the giraffe all the monkey numbers: *gi*, *zo*, *zogi*, and *zozo*. Again, feedback was not given, and the giraffe thanked children for each response.

Giraffe Numbers. After the giraffe thanked them for teaching monkey numbers, children were told that giraffes also have their own numbers that are "a bit like English, and a bit like monkey numbers". They were shown that *one*, *two*, and *three* are the same as English. They were then told that the other giraffe numbers were different, and that they would play the guessing game with some new giraffe numbers. They were then tested on *three-two* (or *three-and-two*), *three-one* (or *three-and-one*), and *three-three* (or *three-and-three*) using the methods of the monkey number task.

Results

For combinations of *gi* and *zo,* children gave correct responses about half of the time in the Explicit Conjunction condition (52.8 ± 7.3%) and about a third of the time in the Implicit condition

(33.9 ± 6.9%; see Figure 3.3). To explore these results, we constructed a mixed-effects linear regression predicting accuracy on the Monkey Numbers task, including as predictors Condition (Explicit vs. Implicit Conjunction), responses to pseudo-English (Giraffe) numbers, their interaction, Target Quantity (4 or 6), Training Trial performance, as well as a random intercept for each child. We found no effect of Condition (b=-1.999, z=-1.402, p=0.161), or of responses to pseudo-English constructions (b=0.202, z=0.156, p=0.876), nor any interaction between them (b=0.473, z=0.288, p=0.773). Thus, we found no evidence that performance on pseudo-English forms was related to novel "monkey" forms. In addition to this, the model also found that children provided correct responses more often for requests of 3+1 than for 3+3 (b=-0.761, z=-4.464, p<0.001), and that accuracy was predicted by performance on training trials (b=10.253, z=3.295, p=0.001). Adding age to this model did not significantly predict performance (b=0.059, z=1.354, p=0.176), possibly because this variance was captured by performance on training trials.

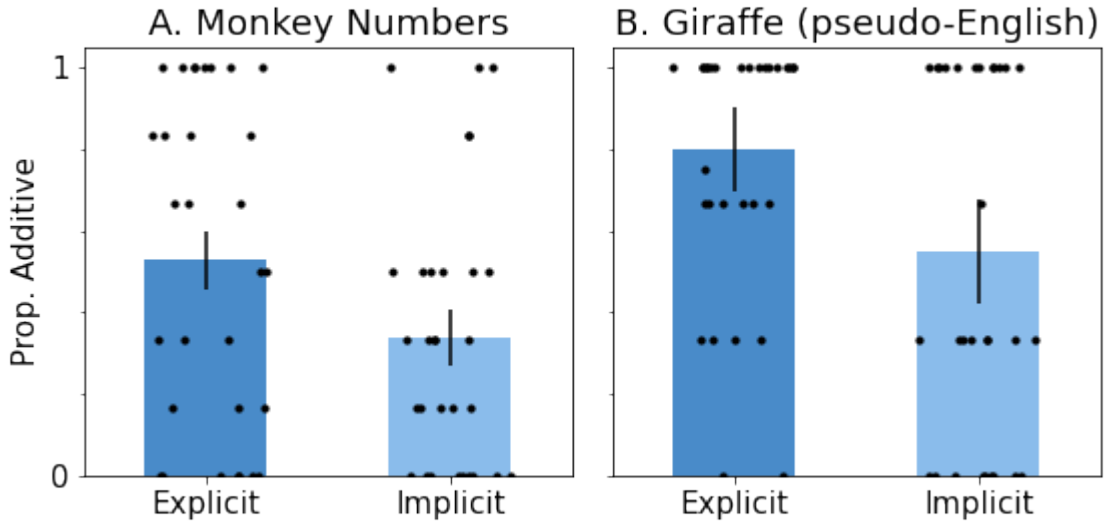


Figure 3.3. A. How often children in each condition (dark: Explicit; light: Implicit) created sets of 3+1 and 3+3 by additively composing each number word in "Monkey numbers". B. Same, for pseudo-English constructions ("Giraffe"). Dots represent individual children.

Children's performance for pseudo-English Giraffe Numbers was higher in both the Explicit (76.9 ± 8.7%) and Implicit (51.1 ± 10.4%) conditions. To explore this difference between number systems, we constructed an exploratory model predicting accuracy from Language (novel Monkey vs. pseudo-English), Condition (Explicit vs. Implicit Conjunction), Target cardinality (4 or 6), Training Trial performance, and a random intercept for each child. This model revealed that responses to pseudo-English numbers were more accurate than to novel Monkey numbers (b=1.739, z=4.484, p<0.001), and that responses in the Explicit condition were more accurate than those in the Implicit condition (b=1.875, z=3.138, p=0.002). Children were also more accurate when responding to requests for 3+1 than to requests of 3+3 (b=0.4, z=3.326, p=0.001), and Training Trial performance again significantly predicted accuracy (b=4.364, z=2.765, p=0.006).

One obvious difference between English numbers and Monkey numbers is that English words are highly familiar, whereas children's memory for novel number words may be more fragile. We therefore wondered how much children's weaker performance for novel words was due to forgetting their meanings versus difficulty in rule learning. To explore this, we examined the last two test trials in which children were asked for just sets of *gi* and of *zo*. We found that roughly ⅓ of children responded incorrectly for both numbers (Implicit: 8; Explicit: 9), though most answered both correctly (Implicit: 19; Explicit: 15), and a few recalled only one (Implicit: 3; Explicit: 6). Of the errors that children made, almost all involved confusing the two number meanings with each other, suggesting that they remembered which cardinalities the words referred to, but forgot which word encoded which cardinality. To explore the impact of memory for *gi* and *zo* on accuracy in understanding combined forms, we replaced "Training Trial performance" in the primary analysis above with children's accuracy on these two trials as a predictor (Memory), as well as its interaction with Target quantity. This model showed that accuracy was lower for

requests of 3+3 than 3+1 (b=-1.1625, z=-3.803, $p<0.001$) compatible with some confusion of *gi* and *zo*. However, it showed no effect of memory for *gi* or *zo* on generalization accuracy (b=-2.524, z=-1.242, $p=0.214$), nor an interaction between memory and target quantity (b=0.634, z=1.684, $p=0.092$). Thus, it appears that learning rules in a novel number system is harder than learning rules for familiar number words, even when controlling for difficulties recalling the meanings of novel words.

As in Experiment 1, we asked what strategies children may have employed (see Table 3.2), but coded children as consistent if they used a strategy on 4/6 trials (since there were only 6 generalization trials). The most common strategy was the Additive strategy (14/19 Explicit; 5/11 Implicit), followed by the First Word strategy (3/19 Explicit; 3/11 Implicit). It is unlikely that children adopted a multiplicative interpretation of composed expressions. If children interpreted requests multiplicatively, we would expect them to sometimes give nine objects in response to requests for *zozo* (as 3x3). However, this never occurred. Similarly, a multiplicative interpretation of *three-two* or *three-and-two* would lead to sets of 6. This happened only once, in the implicit condition.[21]

[21] Children may have had other interpretations not considered here. For example, after giving 4 for *zo and zo*, one 5-year-old responded to a request for *zo and gi* by asking, "Are they four and five?", creating a set of 5, and then responding to the next request with a set of 6.

Table 3.2. The number of children in each condition who were consistent with alternative strategies on at least 4/6 trials.

	Additive	Indexing	First Word	Last Word	Give Many
Explicit	14	1	3	1	0
Implicit	5	1	3	1	1
All	19	2	6	2	1

Discussion

In Experiment 2, we found that CP knowers could learn number word meanings in a novel language, and combine these words to form larger numerical expressions using both explicit and implicit additive rules. Children's ability to understand conjunctive rules for novel words was not predicted by performance for pseudo-English words, and overall was better when rules operated over familiar number words than when they applied to novel numbers.

General Discussion

Why are the compositional rules of number words acquired late in numerical development? In two experiments, we asked whether the ability to interpret combinations of number words using syntactic rules depends on learning the cardinal principle, or if children can learn such rules earlier given training with small numbers. In Experiment 1, we found that children could additively compose small number words prior to acquiring the cardinal principle, but only when this rule was explicitly communicated via the word "and". Cardinal principle knowers also often learn additive rules, and did so even when the additive rule was expressed without explicit "and", and when it

required them to reason about sets greater than 10. Finally, Experiment 1 found that some CP knowers, but fewer subset knowers, learned multiplicative rules for combining numbers. In Experiment 2, we extended these findings by asking whether CP knowers could also learn additive rules in a system anchored at 3, composed entirely of novel words. We found that children were able to interpret additive rules in both systems, though more easily in English. Also, the explicit presence of "and" had only a small effect on performance.

In the Introduction, we presented two possible explanations for why children might acquire rules for combining number words so late in development. On one account, learning the syntax of number words might depend on a prior conceptual understanding of counting. According to several recent accounts, when children first become CP knowers, they learn to accurately label and construct large sets, but lack an understanding that, as one counts up in the count list, the cardinality of the set increases by exactly one (Davidson et al., 2012; Spaepen et al., 2018). Also, interestingly, acquisition of this knowledge has been linked to children's later learning of the syntactic rules of counting (Cheung et al., 2017; Chu et al., 2020; Schneider et al., 2020), both of which emerge between around 5½ to 6 years of age. Given these findings, one possibility is that children can't learn how to compose number words like "twenty" and "one" until (1) they can assign cardinal meanings to each of these numbers (by becoming cardinal principle knowers), and (2) they understand that counting up in the count list involves the addition of 1 item. An alternative explanation, however, is that learning the rules that govern numbers does not depend on a prior conceptual understanding of counting, but is delayed simply because children lack robust exposure to the relevant rule structures. Before the age of 4, most children have limited counting abilities, and often cannot recite numbers beyond 20 or 30. Also, children's language input contains frequent use of small number words like *one* and *two*, but very infrequent use of larger numbers that contain

evidence of syntactic rules (Willits, Jones, & Landy, 2016). Because previously published studies of number word learning feature children learning base-10 systems, naturalistic data from previous studies could not assess this question, leaving open whether rule learning might occur earlier in children exposed to systems composed from smaller numbers. By drawing on conjunctive syntax that is attested in many world languages, and using this syntax to conjoin either familiar or unfamiliar number words, we showed that many children can compose number words relatively early in development, and often prior to acquiring the cardinal principle.

One important limitation to children's performance is that children performed better with additive rules than multiplicative rules, and better when additive rules were expressed explicitly via the conjunctive word "and". In particular, we found that subset knowers, though not CP knowers, were significantly more likely to compose sets additively when number words were combined with the word "and", and that children's responses to multiplicative constructions became more consistent with age. This benefit of explicit cues to conjunction is perhaps not surprising given its presence in many world languages, as described in the Introduction, and also given its possible role in disambiguating the structure of numerical expressions. For example, given a novel expression like "*two-three bananas*" a listener might infer that the speaker intends to refer to 5 items, using an additive interpretation, or to 6, if they assume a multiplicative reading. In fact, in English, exactly such ambiguity exists, given expressions like *sixteen* (six and ten) and *sixty* (six tens).

Another interesting finding of the present study is that, although counting knowledge did not appear to be a prerequisite to learning rules, children with greater counting experience learned rules more easily, especially with increasing age. While the most straightforward interpretation of this result is that children learn information more easily as they grow older, our study leaves open

whether training on counting procedures might indeed facilitate rule learning, above and beyond exposure to the rules themselves. This is because, in our study, CP knowers had knowledge of how to count in English, but no training on counting routines in the trained number systems. This raises the question of whether adding rote training on a count routine composed of trained numbers might facilitate learning, and more generally what relation counting bears to other aspects of a number system (Beller & Bender, 2008). Interestingly, relevant evidence from adult studies of number syntax is mixed (Holt & Barner, 2023). When trained on novel base systems ranging from base-2 to base-10, adult learners are better able to memorize and recall number words when they are presented as part of a rote, sequential, list, but also show poorer comprehension of these numbers, relative to when they are trained on their meanings without exposure to a count list. This raises the possibility that rote counting may make it easier to memorize numbers, but harder to learn the meaning conveyed by their syntactic structure. Future studies should explore this possibility in child learners.

The present study has multiple limitations, leaving open a variety of questions for future work. First, and most importantly, the experiments presented here tested children under limited experimental conditions, leaving open whether training with familiar, larger, numbers might result in earlier learning of attested number system rules. For example, whereas this study shows that, e.g., CP knowers, can spontaneously compose a novel word for 10 with smaller sets using conjunction, future work should investigate whether children can be trained to compose familiar decade labels like *twenty* and *thirty* with unit labels to accelerate their acquisition of these rules within their own language. Second, our study only begins to explore how children learn novel systems. For example, Experiment 2 tests only CP knowers, leaving open whether younger children could as easily learn new numbers and apply rules to them in a single session. Also, our

study tested only select numerical expressions, whereas diverse number systems are attested, presenting many opportunities for studying the affordances of syntactic regularity, conciseness, or ambiguity (Chrisomalis, this volume) on the trajectory of individual learners. For example, Holt and Barner (2023) asked how the affordances of number systems with small or large bases, or those with or without a counting routine, might impact how words and rules are acquired. While their study begins to explore these questions for adults, it's possible that learning proceeds differently for children, or that it differs with more extensive longitudinal training. Lastly, there is considerable evidence that the modality of number systems has important implications for how learners understand number concepts (Chrisomalis, 2004, 2010, 2020; Schlimm, 2018, inter alia). It remains unknown whether children might learn visually presented numerical symbols and rules that govern them more easily than verbal systems, when given training on these systems. Future studies might train children to associate physical tokens or images with small numbers, and test their ability to conjoin these symbols to represent large sets. More generally, the use of artificial number languages offers a new window into how individual learners innovate given the affordances of unfamiliar number systems in a cultural context dominated by base-10.

Acknowledgement

Chapter 3, in full, has been submitted for publication and may appear as Holt, S. & Barner, D. (2025). Children's path to number syntax: concrete logic or procedural abstraction. Under review at *Philosophical Transactions of the Royal Society B.* The dissertation author was the primary researcher and author of this paper.

References

Almoammer, A., Sullivan, J., Donlan, C., Marušič, F., Žaucer, R., O'Donnell, T., & Barner, D. (2013). Grammatical morphology as a source of early number word meanings. *Proceedings of the National Academy of Sciences*, *110*(46), 18448-18453.

Barner, D., Chow, K., & Yang, S. J. (2009). Finding one's meaning: A test of the relation between quantifiers and integers in language development. Cognitive psychology, 58(2), 195-219.

Barner, D., Wood, J., Hauser, M., & Carey, S. (2008). Evidence for a non-linguistic distinction between singular and plural sets in rhesus monkeys. *Cognition, 107*(2), 603–622.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67* (1), 1–48.

Beller, S., & Bender, A. (2008). The limits of counting: Numerical cognition between evolution and culture. *Science*, *319*(5860), 213-215.

Bloom, L., Lahey, M., Hood, L., Lifter, K., & Fiess, K. (1980). Complex sentences: Acquisition of syntactic connectives and the semantic relations they encode. *Journal of Child language, 7*(2), 235-261.

Buijsman, S. (2020). How numerals support new cognitive capacities. *Synthese*, *197(9*), 3779-3796.

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.

Chrisomalis, S. (2004). A cognitive typology for numerical notation. Cambridge Archaeological Journal, 14(1), 37-52.

Chrisomalis, S. (2010). *Numerical notation: A Comparative History*. Cambridge University Press.

Chrisomalis, S. (2020). *Reckonings: Numerals, Cognition, and History*. MIT Press.

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. (2005). Numeral bases. In: The world atlas of language structures. Oxford Univ. Press. p. 530–3.

Cooperrider, K., & Gentner, D. (2019). The career of measurement. *Cognition*, *191*, 103942.

Davidson, K., Eng, K., & Barner, D. (2012). Does learning to count involve a semantic induction? *Cognition, 123*, 162-173.

Donohue, M. (2008). Complexities with restricted numeral systems. *Linguistic Typology, 12*, 423–429.

Epps, P. (2006). Growing a numeral system: The historical development of numerals in an Amazonian language family. *Diachronica*, *23*(2), 259-288.

Epps, P., Bowern, C., Hansen, C. A., Hill, J. H., & Zentz, J. (2012). On numeral complexity in hunter-gatherer languages. *Linguistic Typology, 16,* 41-109.

Falk, R. (2010). The Infinite Challenge: Levels of Conceiving the Endlessness of Numbers. Cognition and Instruction, 28(1), 1–38.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., ... & Stiles, J. (1994). Variability in early communicative development. Monographs of the society for research in child development, i-185.

Gould, P. (2017). Mapping the acquisition of the number word sequence in the first year of school. Mathematics Education Research Journal, 29(1), 93-112.

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. Harrison & Sons.

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. (2023). From partners to populations: A hierarchical Bayesian account of coordination and convention. *Psychological Review, 130*(4), 977.

Holt, S., & Barner, D. (2023). How does the syntax of counting affect learnability? Evidence from artificial language learning. *In Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 45, No. 45).

Hurford, J. R. (1987). Language and number: The emergence of a cognitive system.

Jara-Ettinger, J., Piantadosi, S., Spelke, E. S., Levy, R., & Gibson, E. (2017). Mastery of the logic of natural numbers is not the result of mastery of counting: Evidence from late counters. *Developmental Science, 20(*6), e12459.

Justus, C. F. (1999). Pre-decimal structures in counting and metrology. *Trends in Linguistics. Studies and Monographs, 118*, 55-80.

Le Corre, M., & Carey, S. (2007). One, two, three, four, nothing more: An investigation of the conceptual sources of the verbal counting principles. *Cognition*, 105(2), 395-438.

Le Corre, M., Li, P., Huang, B. H., Jia, G., & Carey, S. (2016). Numerical morphology supports early number word learning: Evidence from a comparison of young Mandarin and English learners. Cognitive psychology, 88, 162-186.

Lust, B., & Mervis, C. A. (1980). Development of coordination in the natural speech of young children. *Journal of Child Language, 7*(2), 279-304.

Marušič, F., Žaucer, R., Plesničar, V., Razboršek, T., Sullivan, J., & Barner, D. (2016). Does grammatical structure accelerate number word learning? Evidence from learners of dual and non-dual dialects of Slovenian. *PloS One*, 11(8), e0159208.

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

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. (1989). Comparisons of US and Japanese first graders' cognitive representation of number and understanding of place value. *Journal of Educational Psychology, 81*(1), 109.

Miura, I. T., Kim, C. C., Chang, C. M., & Okamoto, Y. (1988). Effects of language characteristics on children's cognitive representation of number: Cross-national comparisons. *Child Development*, 1445-1450.

Morris, B. J. (2008). Logically speaking: Evidence for item-based acquisition of the connectives AND & OR. *Journal of Cognition and Development, 9*(1), 67-88.

Pelland, J. C. (2020). What's new: Innovation and enculturation of arithmetical practices. *Synthese, 197(*9), 3797-3822.

Piantadosi, S. T., Jara-Ettinger, J., & Gibson, E. (2014). Children's learning of number words in an indigenous farming-foraging group. Developmental science, 17(4), 553-563.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences, 27*(2), 169-190.

R Core Team. (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing. https://www.R-project.org/

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

Sarnecka, B. W., Kamenskaya, V. G., Yamana, Y., Ogura, T., & Yudovina, Y. B. (2007). From grammatical number to exact numbers: Early meanings of 'one','two', and 'three'in English, Russian, and Japanese. *Cognitive Psychology, 55*(2), 136-168.

Sarnecka, B. W., & Lee, M. D. (2009). Levels of number knowledge during early childhood. Journal of experimental child psychology, 103(3), 325-337.

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.

Schlimm, D. (2018). Numbers through numerals: The constitutive role of external representations. In Naturalizing logico-mathematical knowledge (pp. 195-217). Routledge.

Secada, W. G., Fuson, K. C., & Hall, J. W. (1983). The transition from counting-all to counting-on in addition. *Journal for research in Mathematics Education, 14*(1), 47-57.

Siegler, R. S., & Robinson, M. (1982). The development of numerical understandings. In *Advances in Child Development and Behavior* (Vol. 16, pp. 241-312).

Spaepen, E., Gunderson, E. A., Gibson, D., Goldin-Meadow, S., & Levine, S. C. (2018). Meaning before order: Cardinal principle knowledge predicts improvement in understanding the successor principle and exact ordering. *Cognition, 180*, 59-81.

Spelke, E. S. (2017). Core knowledge, language, and number. *Language Learning and Development*, *13*(2), 147-170.

Sullivan, J., Cramer-Benjamin, S., Alvarez, J., & Barner, D. (2023). Everything is infinite: Children's beliefs about endless space, time, and number. *Open Mind, 7*, 715-731.

Sullivan, J., Alvarez, J., Cramer-Benjamin, S., Holcomb, S., Nolan, M., Morabito, A., Schneider, R. & Barner, D. (under review). Do preschoolers use rules to represent their count list?

Wagner, K., Kimura, K., Cheung, P., & Barner, D. (2015). Why is number word learning hard? Evidence from bilingual learners. *Cognitive Psychology, 83*, 1-21.

Willits, J. A., Jones, M. N., & Landy, D. (2016). Learning that numbers are the same, while learning that they are different. *In Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 38).

Wright, R. J. (1994). A study of the numerical development of 5-year-olds and 6-year-olds. *Educational Studies in Mathematics, 26*(1), 25-44.

Wynn, K., Bloom, P., & Chiang, W. C. (2002). Enumeration of collective entities by 5-month-old infants. *Cognition, 83*(3), B55–B62.

GENERAL DISCUSSION

The systems of symbolic number that have emerged historically in human cultures have enabled people to represent exact quantities far beyond the limits of perceptual mechanisms like the ANS or PI, and to conceive of mathematical abstractions even further removed from direct experience, such as infinity. In addition to helping individuals discover number concepts, symbolic number also constrains numerical thought. In the studies presented above, adults and children reasoned about number in ways that were closely tied to their goals and to formal properties of the symbolic number systems they were using.

In Chapter 1, numerate adult participants relied heavily on 1-to-1 correspondence to communicate about below and above the subitizing limit of 4. This was despite the fact that they were already familiar with sophisticated symbolic number systems, such as the Arabic numerals. In Experiment 2, participants relied on spatial configurations to communicate about much larger numbers. Moreover, they took the opportunity to create more economical, compressed representations by using fewer strokes to represent e.g. three columns of objects rather than every individual object. In Experiment 3, most participants created a cumulative-additive base system by repeating the largest number symbol as often as possible, while others persisted in using 1-to-1 correspondence. Others created alternative syntaxes, such as a pseudo-multiplicative construction for 6, “222”. These findings show that even heavily enculturated adults use a range of strategies for communicating about number, which are thus not completely informed by participants’ familiarity with e.g., Arabic numerals. Importantly, these strategies were deployed in a communicative context. While participants may have considered other ways of expressing numerical information, these are what they relied on to help an anonymous partner understand what to do, and they resemble many of the strategies evidenced in the historical record.

In Chapter 2, other numerate adults learned artificial number systems that featured different arithmetical bases. There was no overall effect of base size, but different correlates of base size were related to participants' ability to learn words and rules. Multiplicative rules were more difficult to learn than additive rules, as were number words that contained multiplicative rules. These results are consistent with participants' having a representation structured by the counting routine, in which number words are stored as a sequence that is rehearsed in canonical order. In Experiment 2, a consistently ordered counting routine helped participants remember number words more effectively than those who learned number words without a counting routine. However, it was learners who had no count routine who performed best when judging the relative magnitudes of number words. This suggests an influence of counting on learners' representations that facilitates memory, perhaps through words' integration into a predictable rhythm, but makes other conceptual less salient, such as cardinal representations of each number.

In Chapter 3, children were introduced to an artificial number system, featuring a small arithmetical base of 2 or 3, and asked to create sets of objects based on numbers in the new system. Children who did not yet know to count four or more objects were able to interpret expressions like "two and two", creating sets consistent with an additive meaning of "and". Further, CP knowers were able to interpret requests for composed number words additively even without the help of "and", but were also helped by the presence of "and" to interpret number words additively. In Experiment 2, CP knowers showed a weaker bias when learning number words from a novel language. They were also more likely to interpret number words additively when composed of familiar English words, rather than novel syllables, suggesting that children may need to develop automatic associations with number words before they are able to reason effectively about compositions of numbers.

One notable bias that participants showed throughout was a preference for cumulative representations similar to those attested in the historical record. For example, many of the adult participants in Chapters 1 and 2 created novel number symbols using cumulative additive rules of composition, and children found number words easier to interpret when "and" signalled a conjunctive operation between familiar numbers. Also, many of participants stumbling blocks were similar to those evidenced in the historical emergence of human symbolic numbers. For example, both adults and children were much less likely to learn or create symbols featuring multiplicative rules of composition. These patterns suggest that we might be able to leverage experiments with modern participants to make inferences about the diachrony of symbolic number. Future work might explore in more depth the different cognitive mechanisms that have been proposed to play a role in the emergence of multiplicative rules. For example, natural language syntax may allow for number words to quantity noun phrases that are already quantified by numbers, such as "two groups of three cats" (Spelke, 2017). While inconclusive, CP knowers' responses to multiplicative requests in Chapter 3 is consistent with having interpreted "two twos of bananas" similarly. Alternatively, other accounts suggest that multiplicative rules are likely to emerge when humans encounter the limits of their finite counting routine, and must express higher multiples by referring to completion of the count routine explicitly (e.g. Hurford, 1987; Damerow, 2007; Dos Santos, 2021).

The experiments here, with dyads and individual learners, set the stage for investigation of how universal cognitive biases influence the ways in which symbolic number systems evolve over time when transmitted between generations of users. While a large body of empirical work on artificial language learning has shed light on the cognitive constraints shaping linguistic typology (see Culbertson, 2023, for review), much less has been directed at the cognitive

constraints on symbolic number. For example, one relevant research strategy in linguistic studies is to identify a typological pattern found across languages, and ask whether it reflects some universal cognitive pressure, or instead is a result of historical contingency. For example, the prevalence of subject-first word order in emerging creoles suggests that it could be the result of universal cognitive constraints. When asked to describe scenes using gestures alone, experimental participants tend to use subject-first order even when their native language is predominantly verb-first (Futrell et al., 2015), consistent with a universal cognitive constraint. Future work could investigate whether the biases shown by participants in the present studies also represent universal cognitive constraints, or are the result of enculturated habits. In Chapter 1, adult participants also showed a strong bias to order symbols in a particular way, often placing the symbols that denote larger quantities earlier in the expression. Would participants show the same bias in another symbolic medium, such as verbal communication? Given that natural language number syntax can show variable word-order in compositional number expressions (e.g. English *twenty-one* versus Dutch *een-en-twintig*, lit. "one and twenty"), future work might explore whether participants' native languages exert a strong influence on their preferences for ordering number words. Unlike natural language, however, participants may be more likely to reflect explicitly on number systems when generalizing their rules beyond the range of what has previously been learned. This type of conscious reflection may, for example, explain the base-10 strategies adopted by some participants in Chapter 2. An important theoretical priority in extending experimental semiotics from natural language to number systems should be to understand how people's conscious inferences about number words lead them to use different strategies from those that that they adopt unconsciously.

The results of Chapter 3 also suggest that studying how children learn novel number systems can help inform how we think about children's acquisition of the cardinal principle. While CP knowledge is tied to counting, even subset knowers were able to create sets beyond their knower level when asked to do so using complex compositions of familiar words. This result suggests that children can interpret complex numerical expressions to create sets much earlier than previously evidenced, and which might in principle change the possible trajectory of number word learning. What these results do not yet tell us, however, is how these children might understand what they have achieved. For example, subset knowers may interpret "two and two" as one complex expression referring to a single number concept. Alternatively, they may interpret "and" in other ways, such as a discourse connective, a meaning with which children may associate "and" before they use it to express conjunctions (Bloom et al., 1980). Such an interpretation is similar to theories mentioned above, that complex number syntax emerges when numbers are used with reference to sequences of actions, rather than to sets of objects.

Besides showing that children can interpret complex number words to create sets before acquiring the cardinal principle, the findings presented here also suggest that counting may help learners in different ways than previously imagined. Specifically, adult participants learning a counting system found it easier to accurately remember words they had learned when tested on them in the counting sequence, but were less aware of the arithmetical relations between number words that they had learned. Together with subset knowers' success at interpreting conjunctive number expressions, these findings suggest that the rote counting procedure may play only an indirect role in learning that numbers concept can be composed. While experience with the counting sequence provides learners with exposure to syntactic rules, it may at first also make learners less likely to reason about the relationship between those rules and the arithmetic

properties that they encode. However, this may ultimately be to learners' advantage. As Dos Santos (2021) notes, describing a theory of the historical emergence of number symbols from 1-to-1 correspondence, "*automatization of the procedure is crucial for its successful execution, since it prevents distraction. If at each step the operator stopped to grasp how many fingers she had already used, she could lose track of the tally. The procedure is more accurate if executed uninterruptedly*." It may be that before gaining a critical amount of experience with accurately counting large quantities, young CP knowers' attempts to reflect on the cardinal meaning of the counting list and its number syntax would be undermined by their inability to consistently create sets that reflect these meanings. If that is the case, then children's acquisition of abstract number knowledge may be best served by learning the structure of the counting procedure without reasoning about its cardinal entailments. If so, an adult-like understanding of the arithmetical properties of the count list may emerge most efficiently when children have gained familiarity with large portions of the count list, and can reflect on its structure as a whole. Alternatively, perhaps learning a base system in which subitizable numbers are composed with each other could make the arithmetical properties of number words salient to children even before they have learned how to count. A third possibility is that reasoning about the count routine separately from arithmetic may be useful for children, but not for adults. For example, children are often more sensitive than adults to superficial properties of language (Culbertson & Schuler, 2019) and other symbolic systems, and introduce systematicity into artificial languages without necessarily introducing or understanding compositional structure (Raviv & Arnon, 2018). They may therefore benefit from learning the entire count routine prior to reflecting on its meaning, in a way that is not cognitively useful for adults. Future work could develop the methods used here to test these questions directly.

Finally, studying how participants from numerate cultures reason about artificial number systems can help provide new insights into analogies that are often drawn between the ontogenic and diachronic emergence of numerical abstractions. For example, Piaget (1928) thought about stages in children's cognitive development as analogous to epochs in the history of mathematics, and considered that both ontogeny and diachrony may reflect universal constraints on human cognition, which operate both within the minds of individuals and outside of the head, in human culture. While there is a widely acknowledged correlation between sociocultural complexity and numerical abstraction, attempts to account for specific causal mechanisms linking the two, such as subsistence patterns (Epps et al., 2012) or climatic conditions (Divale, 1999), have not been conclusive (O'Shaughnessy, Gibson, & Piantadosi, 2022). Future experimental work using methods like those presented here could target specific hypotheses about how properties of social networks influence the emergence of abstract symbolic number in human culture. For example, many prenumerate societies begin to use symbolic number only when they become economically integrated with numerate societies, and individuals' reasoning about number in these cultures is often closely tied to economic activities like trading (e.g., O'Shaughnessy et al., 2023). While the experiments presented here involved creating and learning symbolic systems, the social goals that participants had to achieve were relatively narrow. However, the methods used here can be modified to simulate the kinds of complicated social interactions within which number symbols are thought to have historically emerged, such as exchanges or timekeeping. For example, participants in Chapter 1 had to cooperate with each other in order to identify target arrays from distractors. However, they may have converged on different strategies if each member of the dyad also had to achieve a goal that conflicted with their partner's.

REFERENCES

Culbertson, J., & Schuler, K. (2019). Artificial language learning in children. *Annual Review of Linguistics, 5*(1), 353-373.

Culbertson, J. (2023). Artificial language learning. *Oxford Handbook of Experimental Syntax*, 271-301.

Damerow, P. (1996). Number as a second-order concept. *Science in context, 9*(2), 139-149.

Damerow, P. (2007). The material culture of calculation: A theoretical framework for a historical epistemology of the concept of number. In *Mathematisation and Demathematisation* (pp. 19-56). Brill.

Divale, W. (1999). Climatic instability, food storage, and the development of numerical counting: A cross-cultural study. *Cross-Cultural Research, 33*(4), 341-368.

Dos Santos, C. F. (2021). Enculturation and the historical origins of number words and concepts. *Synthese, 199*(3), 9257-9287.

Epps, P., Bowern, C., Hansen, C. A., Hill, J. H., & Zentz, J. (2012). On numeral complexity in hunter-gatherer languages. *Linguistic Typology, 16,* 41-109.

Futrell, R., Hickey, T., Lee, A., Lim, E., Luchkina, E., & Gibson, E. (2015). Cross-linguistic gestures reflect typological universals: A subject-initial, verb-final bias in speakers of diverse languages. *Cognition, 136*, 215-221.

Hurford, J. R. (1987). Language and Number: The emergence of a cognitive system. *Oxford: Basil Blackwell*.

O'Shaughnessy, D. M., Gibson, E., & Piantadosi, S. T. (2022). The cultural origins of symbolic number. *Psychological Review, 129*(6), 1442.

O'Shaughnessy, D. M., Cruz Cordero, T., Mollica, F., Boni, I., Jara-Ettinger, J., Gibson, E., & Piantadosi, S. T. (2023). Diverse mathematical knowledge among indigenous Amazonians. *Proceedings of the National Academy of Sciences, 120*(35), e2215999120.

Piaget, J. (1928). Logique génétique et sociologie. *Revue Philosophique de la France et de l'Étranger, 105*, 167-205.

Raviv, L., & Arnon, I. (2018). Systematicity, but not compositionality: Examining the emergence of linguistic structure in children and adults using iterated learning. *Cognition, 181*, 160-173.

Spelke, E. S. (2017). Core knowledge, language, and number. *Language Learning and Development, 13*(2), 147-170.