

UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Improvised Numerals Rely on 1-to-1 Correspondence

Permalink

<https://escholarship.org/uc/item/12m8w2gk>

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 43(43)

ISSN

1069-7977

Authors

Holt, Sebastian

Barner, David

Fan, Judith E.

Publication Date

2021

Peer reviewed

Improvised Numerals Rely on 1-to-1 Correspondence

Sebastian Holt, David Barner, Judith E. Fan

Department of Psychology, University of California San Diego

{sholt, dbarner, jefan}@ucsd.edu

Abstract

Symbolic representations of number are instrumental to mathematical reasoning and many aspects of social organization. What explains their emergence in human cultures? To understand how functional and cognitive constraints impact people's communication about number, we used a drawing-based reference game to investigate how human dyads coordinated to form novel number systems. We found a systematic bias towards symbols exploiting 1-to-1 correspondence to objects in visual arrays, and that this strategy was contingent on the communicative relevance of number. Moreover, the meaning of these symbols was transparent to third party observers not present during their production. Finally, model-based analyses of these symbols' visual properties suggest that the ability to decode exact quantity from them may rely on perceptual processing mechanisms beyond those sufficient for object recognition. These findings contribute to our understanding of how both communicative need and capacity for visual abstraction constrain the emergence of iconic representations of exact number.

Keywords: number; numerals; communication; iconicity

Introduction

Numerals are among the most powerful technologies behind humans' capacity for complex social organization. They have emerged repeatedly in a range of distinct forms over human history in tandem with the emergence of state societies and have acted in different ways to both promote the economic interactions of groups and reinforce the power and prestige of the state (Chrisomalis, 2020). Simultaneously, numerical notation has served as a potent computational tool for engineering, timekeeping, and all manner of scientific inquiry.

Numerals are a pervasive and powerful tool, but the ability to wield them is not trivial: humans must overcome bandwidth limitations in both perception (Jevons, 1871; Dehaene & Cohen, 1994) and working memory (Cowan, 2010; Baddeley, 1986) to encode large quantities accurately. Moreover, many important numerical concepts are not directly observable (e.g. negative numbers) or are too large to be easily comprehended (e.g. a billion).

The historical diversity of numeral systems provides clues about how we overcome these challenges, and raises the question of what cognitive mechanisms and functional constraints might explain their origins. A common pattern in numeral systems is the use of iconic, 1-to-1 matching strategies to compose small numbers, and non-iconic conventions to encode larger numbers (Ifrah, 2000). For example, a Roman numeral like 'II' is iconic: it bears a structural similarity to its meaning (one stroke for each member of any

set of 'two'), and its meaning can be derived perceptually at a glance. Numerals that exclusively use 1-to-1 matching are frequently attested in newly emergent numeral systems, such as those documented in ancient Mesopotamia (Schmandt-Besserat, 1980). However, this convention produces problems beyond the *subitizing limit*: 'IIIIIIIIII' cannot be quickly deciphered. To avoid this issue, many (but not all) numeral systems feature non-iconic symbols: the numeral 'XII' represents the same quantity, and can also be read at a glance – though only by those who already know the arbitrary convention. What conditions elicit – or preclude – the use of such non-iconic numerals, and the complex rules of composition that govern them (e.g., why must the 'II' come after the 'X' and not before?)?

Recent laboratory studies suggest the importance of iconicity for constraining the tokens that emerge in ad hoc communication systems. Studies employing drawing (Hawkins, Sano, Goodman, & Fan, 2019; Mukherjee, Hawkins, & Fan, 2019), toy languages with written nonsense syllables (Raviv, Meyer, & Lev-Ari, 2019), and whistled languages (Hofer & Levy, 2019; Verhoef, Kirby, & Padden, 2011; Verhoef, Kirby, & de Boer, 2016) all indicate that communicative systems often feature some iconicity in their early stages, but that this iconicity is often sacrificed in favour of combinatorial syntactic rules (Hofer & Levy, 2019) or abstracted social conventions (Hawkins et al., 2019). This work also shows that humans can successfully isolate functionally relevant information in novel communication systems (Winters, Kirby, & Smith, 2015). This capacity for accommodating situational context is especially relevant for early numerals, which often feature qualitative information about their referent (the thing being quantified; e.g. a numeral for '3' might be specific to goats, or grain), and only gradually isolate quantitative information (i.e. fully general, abstract numerals). Humans' strong, shared inductive biases for efficient combinatorial systems and economy of expression frequently produce non-iconic shorthands that map to their meaning by conventions shared by participants rather than by iconic resemblance.

Do communicative constraints also favor the adoption of non-iconic representations in the domain of number? For representing shapes, iconicity is often powerful and succinct: the meaning of an image that looks like its referent can be discerned immediately due to low-level perceptual similarity (e.g. a photo of a famous person). By some accounts, numerosity is also a low-level, independent primary visual

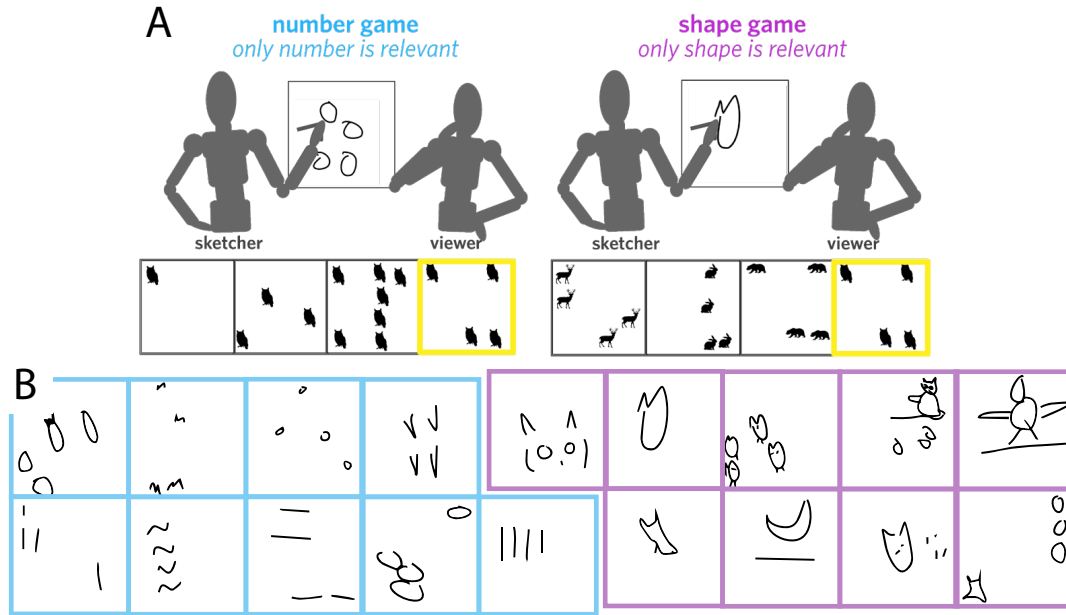


Figure 1: A. In each type of game (number or shape), the sketcher produced a symbol to convey the meaning of the target image (highlighted in yellow) to the viewer, who could see the same sets of animals but not the identity of the target. The sets varied in the type of animal they consisted of (in shape games), or the quantity of that animal they consisted of (in number games). B. Sketches produced by participants in nine different number games (left) and nine different shape games (right) to convey stimulus image “four owls” in the context of number and shape distractors, respectively.

property (Burr & Ross, 2008), which may be especially well represented by iconic symbols that exploit one-to-one correspondence. However, while small sets can be readily represented by one-to-one correspondence, this strategy may be less efficient for sets beyond the subitizing limit, as noise accumulates in keeping with Weber’s law. Thus, we might expect that, when seeking to accurately communicate the values of large sets, individuals may create conventional, non-iconic, representations as they do in other communicative contexts.

In the present study, we asked whether the above trade-offs between iconicity and social convention might also explain the forms that numeral systems have taken historically, by asking what forms first emerge in real-time communicative settings. To explore this, we first constructed a dyadic communication game to elicit spontaneous graphical representations of numerical or shape information. Having elicited these graphical representations under moderate time pressure, we then asked naïve human recognizers to recover the numerical or shape meaning of those representations, in order to understand whether those representations were iconic (and therefore interpretable to naïve outsiders) or conventional (and therefore not interpretable to naïve outsiders). We also fed the image representations into a convolutional neural network trained on naturalistic image data, and gave the same recognition task to a linear classifier trained on those neural-network encodings, in order to see if a low-level measure of perceptual similarity was responsible for observed iconicity in the numeral or shape representations.

What Numerals do People Improvise?

In Experiment I, we elicited ad-hoc numerals from people by putting them in situations where the encoding of numerical information was task relevant. We then compared the symbols produced by this ‘number group’ with symbols produced by a ‘shape group’, another group for whom shape - not numerical - information was task relevant.

Methods

Participants We originally recruited 134 participants from Amazon Mechanical Turk (AMT), paired into 67 games. In each game, one participant was assigned the sketcher role and the other the viewer role, and these role assignments remained the same throughout the experiment. Six games were excluded according to our preregistered criteria: four did not meet the performance threshold of 50% accuracy, while two others included pre-existing symbols (words in the form of text). This left 61 games for analysis, including 122 participants, of whom 115 completed our optional demographic survey (53 female). Each game was either a number game (32 games) or shape game (29 games). All participants provided informed consent as per the IRB.

Stimuli Each stimulus consisted of a set of identical black animal silhouettes on a white, square background (see Figure 1, panel A). Features that varied between stimuli were cardinality (how many of the animal there was in the set) and shape (what kind of animal the set consisted of). A stimu-

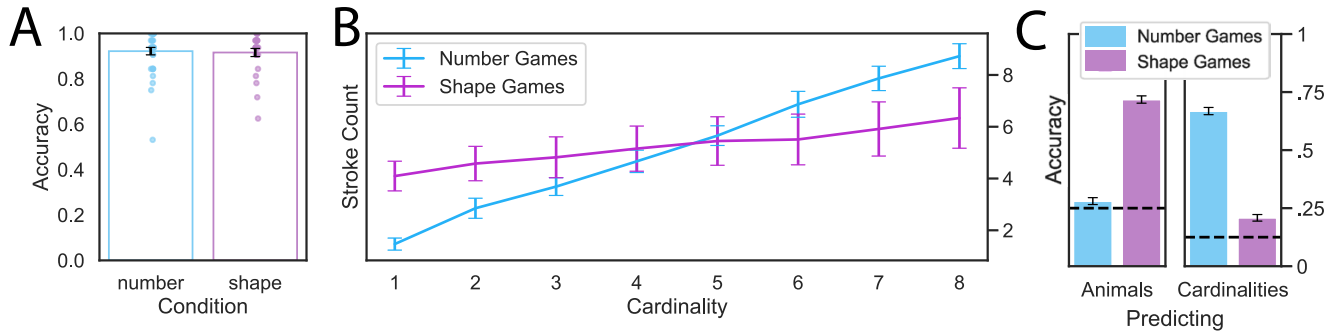


Figure 2: A. Accuracy in the production task for each condition was 92%; individual participant accuracy scores are indicated as dots, with 95% confidence intervals in black. B. The number of strokes used in each sketch correlated much more strongly with cardinality of the stimulus in the number condition than in the shape condition. C. Human recognizer accuracy for labelling sketches produced in the number condition (blue) and shape condition (magenta) with shape (left) or number (right) labels (chance in dotted lines & 95% CI in black).

lus might therefore contain three bears, five deer, five rabbits, etc. There were four possible animals (bears, deer, owls, & rabbits) and cardinalities ranging from 1-8, for a total of 32 unique stimuli. The animal silhouettes were scattered in a random spatial arrangement on the white background, and every time a stimulus was displayed, one of 100 randomly generated versions was selected. To discourage reliance on spatial information, participants were informed that they would see a different spatial distribution from their partner.

Procedure Every game consisted of 32 trials, in which a sketcher had to make symbols that were interpreted by a viewer. In each trial, both participants were shown the same four sets of animals as each other. From those four stimuli, one target was highlighted to the sketcher among three distractors. The job of the sketcher was to disambiguate the target to the viewer by means of a 500x500px sketchpad on which they could draw with their computer cursor. After the sketcher created and submitted their sketch, the viewer saw the sketch and had to guess the identity of the target. This whole process (sketching + guessing) was constrained to 30 seconds per trial, with a progress bar at the top showing both participants how much time remained. At the end of each trial, both participants were shown the identity of the target and the viewer's guess. Because there were 32 trials total per game, each unique stimulus (combination of one of the four animals with one of the eight cardinalities) served as the target exactly once. The 32 trials were divided into four blocks of eight, during which each cardinality and each animal appeared. In this way, we could probe whether representations of each cardinality and each animal changed over time. This also allowed us to create a dataset from each game in which each of the 32 unique stimuli were represented once, enabling easy comparisons based on unique stimuli.

Results

Information is effectively encoded To estimate the factors influencing task accuracy, we constructed a binomial linear mixed-effects model with fixed effects as follows: the shape of the animal in the target (bear, deer, owl, rabbit), cardinality of the target (1-8), condition of the game in which the sketch was produced (shape or number), and block of trials (32 trials / 8 trials per block = 4 blocks), as well as random intercepts for each participant. Through multiple BIC comparisons we found that the best model to predict accuracy did not include any interaction terms. We first asked whether viewers' accuracy in the number condition was different from our baseline shape condition. Overall accuracy was quite high (92%), and it was not significantly different between conditions ($b=-0.221$, $z=-0.623$, $p=.534$), showing that people can successfully distill shape or numerical information when it is relevant in the communicative context (Figure 2, panel A). Regardless of condition, accuracy improved over the the four blocks of trials, indicating that participants did not immediately plateau to their maximum accuracy ($b=0.494$, $z=5.745$, $p<.001$). We also asked whether accuracy was sensitive to the dimensions of variation in the stimuli: cardinality and shape. There were modest differences in accuracy between shape categories, as revealed by nested model comparison ($\chi^2_{(3)}=9.545$, $p=0.023$). By contrast, cardinality was a strong predictor of accuracy ($\chi^2_{(1)}=39.227$, $p<.001$). Higher cardinality was anti-correlated with accuracy ($b=-0.252$, $z=-6.001$, $p<.001$), showing that larger cardinalities were harder for participants to encode and interpret. This may be driven by the perceptual difficulty of number discrimination as cardinalities exceed the subitizing limit.

Strokes are used in 1-to-1 correspondence with numerical meaning We then examined the number of strokes used per sketch. In aggregate, this measure did not differ between the

two conditions; however, accounting for the random effects of participant ID, shape games used more strokes than number games ($b=1.688$, $t=3.103$, $p=0.003$). Participants used fewer strokes with each block of trials ($b=-0.553$, $t=-11.029$, $p<.001$), perhaps as the result of converging on more efficient representations. Importantly, there was both a main effect of stimulus cardinality on the number of strokes used in a sketch ($b=1.0287$, $t=15.782$, $p<.001$), and an interaction between cardinality and game condition whereby stroke counts in number games had a much stronger correlation with stimulus cardinality than in shape games ($b=0.734$, $t=7.767$, $p<.001$) (see Figure 2, panel B). This suggests that sketchers relied heavily on a 1-to-1 matching strategy in the creation of their numerals, to a degree that shape-game sketchers did not (for intuition, see sample sketches in Figure 1, panel B)).

Sketches become more efficient over time Both the amount of time and 'ink' taken to draw each sketch were greater for the shape condition than the number condition (time: $b=5.45$, $t=6.741$, $p<.001$; 'ink': $b=0.0336$, $t=8.98$, $p<.001$). This may imply that the shape stimuli used in our study were slightly more difficult to draw and less efficient than the set of cardinalities. There was also a negative effect of trial block number on the time it took to draw ($b=-1.369$, $t=-23.051$, $p<.001$), and a significant but small effect on the total amount of 'ink' used ($b=-0.0015$, $t=-4.742$, $p<.001$), indicating that participants became more efficient with practice.

How Iconic are Improvised Numerals?

Having elicited improvised numerals, we asked in a follow-up experiment to what extent these numerals would be intelligible to observers who were not part of the original communicative paradigm. If they were, this would indicate that the symbols were iconic; otherwise, we would infer the symbols were formed using non-iconic conventions. Also, to explore the visual processing mechanisms that support the production and recognition of such symbols, we sought to test how well a modern computer vision algorithm that had already achieved high accuracy on other visual recognition tasks (though not number recognition) would generalize to the set of symbols that participants had produced.

Methods

Participants We recruited a total of 112 recognizers from Amazon Mechanical Turk. Six recognizers were excluded according to preregistered criterion of missing any one of four catch trials, and nine further participants were excluded for having participated in the production task, leaving 99 recognizers of whom 95 completed our optional demographic survey (39 female; $M_{age} = 39.02$ years, $SD_{age} = 11.40$ years,). All participants provided informed consent as per the IRB.

Procedure Recognizers were divided into two groups: a number group, in which participants assigned numerical labels to sketches, and a shape group, in which participants assigned animal labels to sketches. Each recognizer saw exactly one randomly selected sketch from each original production

game, of which there were 61. This meant that each recognizer (both 'number-' and 'shape-recognizers') provided labels both for sketches produced in number games and for sketches produced in shape games. 4 catch trials were inserted at regular intervals to exclude recognizers who were not paying attention, for a total of 65 trials. In each trial, the recognizer was presented with a sketch, and also presented with a panel of labelling buttons. For the number group, there were 8 buttons, showing the numbers 1-8. For the shape group, there were 4 buttons, showing the four animal silhouettes used for the original production task stimuli. Recognizers were asked to select the label that best 'matched' the sketch. After selecting their response, no feedback was given.

Model-Based Analysis of Sketch Recognizability To supplement the decisions of human recognizers, we also asked whether the representations encoded by a deep convolutional neural network trained on challenging object recognition tasks would be sufficient to emulate human performance. Following prior work employing deep convolutional neural networks pre-trained on challenging visual recognition tasks (Deng et al., 2009) to encode the visual features of hand-drawn images (Fan, Hawkins, Wu, & Goodman, 2020), we used VGG-19 (Simonyan & Zisserman, 2014) to extract high-level feature-vector representations for each symbol. We trained two logistic-regression classifiers, one predicting number (i.e., 1-8) and the second predicting shape identity from these feature vectors. Both classifiers were trained and evaluated in a 5-fold cross-validated manner using the same splits for both the number and shape tasks, yielding a pattern of correct classifications and confusions.

Results

Relevant information is encoded iconically To measure the degree to which information about shape or number was encoded in a way that was transparent to other people, we analyzed both human and model recognition performance. Specifically, we separately fit human and model recognition accuracy using mixed-effects logistic regression. To identify the factors that were most important for explaining recognition accuracy, we formally compared models containing different subsets of predictors as fixed effects, and used BIC (Bayesian information criterion) to identify the best-performing one. The resulting model predicting human recognition included: production-game condition (i.e., number or shape), recognition-task condition (i.e., number or shape), production-game block number (i.e., 1-4), and the interaction between production-game condition and recognition-task condition, with random intercepts for each recognizer, each production game, and for each stimulus cardinality and shape. Recognition accuracy was slightly lower for sketches produced in shape games than those from number games ($b=-2.365$, $z=-23.775$, $p<.001$), and for recognition participants trying to assign shape labels ($b=-1.738$, $z=-20.143$, $p<.001$). Crucially, however this analysis also revealed a reliable interaction between communication-game

condition and recognition-task condition ($b=4.368$, $z=50.565$, $p<.001$; Figure 2, panel C), indicating that recognition participants were better able to correctly identify the number conveyed by sketches from number games than from shape ones (66.9%, 95CI=[65.3%,68.4%]), and to correctly identify the shape conveyed by sketches from shape games than number ones (71.8%, 95CI=[70.2%,73.3%]). Furthermore, recognition of the feature *not* elicited in the original production game was close to chance-level performance (number labels from shape games 20.8%, 95CI=[19.4%,22.2%]; shape labels from number games 28.0%, 95CI=[26.6%,29.5%]). This suggests that sketchers were selectively prioritizing task-relevant information in their sketches (see Figure 3 for full accuracy/confusion matrices of human recognizers). Like the human recognition participants, the computer-vision model performed better at assigning labels that were congruent with the task-relevant information in the production condition, as indicated by an interaction between production condition and recognition condition ($b=3.338$, $z=20.984$, $p<.001$). However, while its performance classifying shape from shape-game sketches (76.5%, 95CI=[73.6%,79.4%]) was comparable to human recognizers' performance and its failures were at chance (shape labels from number games 25.5%, 95CI=[23.3%,27.7%]; number labels from shape games 14.3%, 95CI=[11.6%,17.1%]), the linear classifier was much worse at discriminating numerical information from number-game sketches (31.7%, 95CI=[30.3%,33.2%]) than were human recognizers. Taken together, these results are consistent with prior work that has shown that similar neural network architectures, previously trained to categorize objects in photos, display a robust ability to generalize to object categorization in sketches (Fan, Yamins, & Turk-Browne, 2018). However, our results also suggest the possibility that the ability to acquire other abstract visual concepts, such as exact cardinality, may not readily emerge under the same architectural and functional constraints, and thus that other computational ingredients may be necessary to support the full breadth of human-like image and symbol understanding.

Discussion

In this study, we began an exploration of how communicative goals and perceptual constraints would affect how people symbolically represent number. We found that numerical and shape information were both encoded iconically by communicators when it was task relevant, as evidenced by the high performance of naïve recognizers. Furthermore, task-relevant information was encoded more iconically than similarly salient, but non-task-relevant information, as evidenced by the relative success of naïve recognizers in recovering task-relevant information and poor performance in recovering non-task-relevant information. While this shows, unsurprisingly, that people are sensitive to the number/shape distinction, more interesting is that it also shows that they create symbols that are restricted to conveying number, rather than also incorporating object information (as happens in many

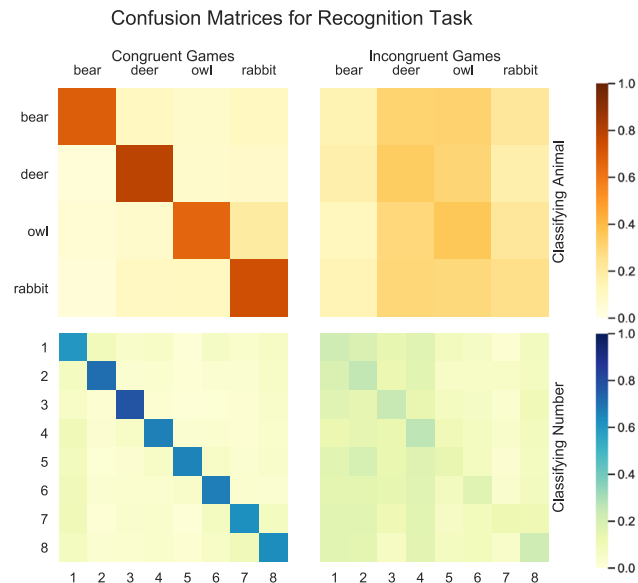


Figure 3: Confusion matrices from decoding. Human recognizers were able to consistently decode sketches produced in conditions congruent to their own (left column of panels), showing effective distillation of the relevant dimension of information by sketchers in the production task. They were much less effective at decoding information from conditions incongruent to their own (right column of panels).

early numeral systems, as well as linguistic number classifiers). This is concordant with previous findings in the field showing that people can (and do) isolate functionally relevant features in communicative signals (Winters et al., 2015).

In addition to iconicity and communicative efficacy, we also wondered what specific strategies people would use to represent numerical information. It was possible that people would be extremely conservative, depicting the stimulus images literally. It was also possible that people would automatically implement an efficient, abstract numerical base system (such as Roman 'V'). This seemed especially likely given the perceptual constraints in the upper range of the cardinalities we tested, and the fact that our participants were certain to already be familiar with the concept of non-iconic numerical symbols. However in our study, somewhat surprisingly, the numerical representations that were elicited relied almost entirely on 1-to-1 iconicity. These representations were not completely conservative, in that they shed extraneous information about shape, and not totally efficient, in that they continued to use 1-to-1 matching for large, perceptually challenging quantities.

This preliminary study introduced a novel approach to investigating how humans create conventions for representing symbolic number. However, the procedure we used has several limitations. One is the small range of cardinalities and

nonrestrictive constraints on numeral production. Currently, we are using larger cardinalities to observe whether base-systems emerge when there is sufficient perceptual noise in numeral representations, and when there are enough costs (limited time and 'ink' resources) to preserving 1-to-1 iconic conventions. Another limitation is that, as they were recruited from Amazon Mechanical Turk, participants can be assumed to be familiar with symbolic number already. In the future, we therefore plan to conduct similar, simplified studies with both pre-numerate child populations, and non-numerate adult populations. Such experiments will be especially informative in light of recent work that finds 1-to-1 matching to be both culturally non-universal and developmentally challenging (Ferrigno, Jara-Ettinger, Piantadosi, & Cantlon, 2017; Schneider & Barner, 2020). These findings suggest that the 1-to-1 matching strategy used by most participants in the present study may not be intuitive to all populations. Finally, our paradigm was not a repeated reference game and featured only a small number of trials (32 trials). By extending the amount of time that participants spend communicating about the same stimuli, we expect that future studies may be more likely to elicit measurable change in the form of representations over the course of the experiment.

OSF pre-registration:

<https://osf.io/y2f3e/>

All code and materials available at:

https://github.com/cogtoolslab/iterated_number_cogsci2021

Acknowledgements

Many thanks to members of the Language Development Lab, and Cognitive Tools Lab at UC San Diego for insightful feedback throughout the process. This work was supported by NSF CAREER Award 2047191 to J.E.F.

References

Baddeley, A. (1986). *Working memory clarendon press*. Oxford.

Burr, D., & Ross, J. (2008). A visual sense of number. *Current biology*, 18(6), 425–428.

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

Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why? *Current directions in psychological science*, 19(1), 51–57.

Dehaene, S., & Cohen, L. (1994). Dissociable mechanisms of subitizing and counting: Neuropsychological evidence from simultanagnosic patients. *Journal of Experimental Psychology: Human Perception and Performance*, 20(5), 958.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition* (pp. 248–255).

Fan, J. E., Hawkins, R. D., Wu, M., & Goodman, N. D. (2020). Pragmatic inference and visual abstraction enable contextual flexibility during visual communication. *Computational Brain & Behavior*, 3(1), 86–101.

Fan, J. E., Yamins, D. L., & Turk-Browne, N. B. (2018). Common object representations for visual production and recognition. *Cognitive science*, 42(8), 2670–2698.

Ferrigno, S., Jara-Ettinger, J., Piantadosi, S. T., & Cantlon, J. F. (2017). Universal and uniquely human factors in spontaneous number perception. *Nature communications*, 8(1), 1–10.

Hawkins, R. X., Sano, M., Goodman, N. D., & Fan, J. W. (2019). Disentangling contributions of visual information and interaction history in the formation of graphical conventions. In *Annual conference of the cognitive science society*.

Hofer, M., & Levy, R. P. (2019). Iconicity and structure in the emergence of combinatoriality. In *Annual conference of the cognitive science society*.

Ifrah, G. (2000). *The universal history of numbers*. Harvill London.

Jevons, W. S. (1871). *The power of numerical discrimination*. Nature Publishing Group.

Mukherjee, K., Hawkins, R. X., & Fan, J. W. (2019). Communicating semantic part information in drawings. In *Cogsci* (pp. 2413–2419).

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

Schmandt-Besserat, D. (1980). The envelopes that bear the first writing. *Technology and Culture*, 21(3), 357–385.

Schneider, R. M., & Barner, D. (2020). Children use one-to-one correspondence to establish equality after learning to count.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

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

Verhoef, T., Kirby, S., & Padden, C. (2011). Cultural emergence of combinatorial structure in an artificial whistled language. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 33).

Winters, J., Kirby, S., & Smith, K. (2015). Languages adapt to their contextual niche. *Language and Cognition*, 7(3), 415–449.