Title
Iconic Artificial Language Learning: A Conceptual Replication with English Speakers

Permalink
https://escholarship.org/uc/item/7b66s7c6

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 45(45)

Authors
Steinert-Threlkeld, Shane
Tachikawa Shapiro, Naomi

Publication Date
2023

Peer reviewed

Iconic Artificial Language Learning: A Conceptual Replication with English Speakers

Naomi Tachikawa Shapiro (tsnaomi@uw.edu)
Department of Linguistics, University of Washington, Seattle, WA 98195 USA

Shane Steinert-Threlkeld (shanest@uw.edu)
Department of Linguistics, University of Washington, Seattle, WA 98195 USA

Abstract

We introduce an *iconic* approach to artificial language learning, one that replaces traditional phonologically-grounded stimuli with pictographic writing systems. Conducting an experiment with English speakers, we demonstrate the viability of this approach by reproducing word order effects observed in multiple studies (e.g., Culbertson & Adger, 2014; Martin, Holtz, Abels, Adger, & Culbertson, 2020). Importantly, iconic artificial languages make it possible to re-use the same linguistic stimuli with diverse language populations, facilitating crosslinguistic investigations.

Keywords: artificial language learning; iconicity; pictograms; word order; semantic scope; typology; language universals

Introduction

Artificial languages have emerged as an insightful tool for probing language learning and processing. In the artificial language learning (ALL) paradigm, participants are taught miniature constructed languages in controlled settings to reveal properties of the human linguistic system. These investigations have explored topics fundamental to language, such as communicative efficiency, statistical learning, hypothesized cognitive constraints on language learning, the origin of crosslinguistic patterns, and the complicated relationships therein. (See Fedzechkina, Newport, & Florian Jaeger, 2016; Culbertson, 2023, for reviews.)

With the exception of gesture-based studies (e.g., Futrell et al., 2015), ALL experiments traditionally employ artificial languages that are grounded in the native phonotactics of the participating speakers, making them easier for the participants to learn but also limiting the usability of the stimuli across a diverse speaker pool. Consequently, most ALL studies focus on a single language population, putting them at risk of overlooking (i) transfer effects from their participants' primary languages that may modulate their findings and (ii) crosslinguistic variation in how statistical universals of language can surface.

In this paper, we thus set out to design linguistic stimuli that strive toward "language neutrality". In particular, we propose rooting artificial languages in *iconic lexicons*, replacing phonologically-realized stimuli with miniature pictographic writing systems, thereby enabling ALL scholars to work with a greater diversity of language populations. In principle, the same neutral artificial language could be taught to participants from typologically diverse linguistic communities, facilitating crosslinguistic analyses and, by extension, the study of transfer effects—and with finer experimental control. This, in turn, could broaden language representation in the ALL literature. Likewise, pictographic languages can be used with adults and children, speakers and signers, as well as monolingual, bilingual, and multilingual populations.

To date, iconicity and non-orthographic symbols have been leveraged to a limited degree in ALL work. In a recent sign language learning experiment, Sato, Schouwstra, Flaherty, and Kirby (2020) showed that iconic gestures helped participants learn form-meaning mappings compared to non-iconic gestures.[1] In the written modality, Saratsli, Bartell, and Papafragou (2020) evidenced preferential marking of certain sources of information over others, even when substituting a nonce evidentiality morpheme with an arbitrary symbol (a black-filled circle). Martin and Culbertson (2020) and Culbertson and Kirby (2022) both used geometric symbols with no associated meanings to study possible domain-general biases for properties of morpheme order. Bowerman and Smith (2022) used iconic stimuli in an iterated communication game to study semantic extension. Finally, though not conceived of as an ALL study, Dautriche, Buccola, Berthet, Fagot, and Chemla (2022) taught simple iconic symbols to *Papio papio* baboons, finding evidence that baboons can entertain compositional representations, namely, negation structures.

To our knowledge, the present ALL study is the first to utilize a fully iconic artificial language in the written modality (with *humans*, that is), where the expressions have associated meanings. We test the viability of iconic stimuli by using a miniature pictographic language to replicate prior ALL findings on linear word order in noun phrases (Culbertson & Adger, 2014; Martin, Ratitamkul, Abels, Adger, & Culbertson, 2019; Martin et al., 2020). We describe these studies in the following section before our experiment. Crucially, we conducted our experiment with English-speaking participants, replicating word order effects that have been attested with English speakers in the aforementioned literature. This replication represents a proof-of-concept that fully iconic stimuli can be used successfully in ALL research. Future work should use iconic stimuli to conduct ALL studies with

[1] Indeed, some research has argued that iconicity is an essential ingredient for language development, providing scaffolding that enables learners to link embodied experience to linguistic form and, ultimately, to a symbolic system that is mostly arbitrary. For discussion, see Perniss, Thompson, and Vigliocco (2010).

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

linguistically diverse (non-English-speaking) populations—a point we return to in the Discussion.

Scope-Isomorphism in the Noun Phrase

Greenberg's (1963) *Universal 20* noted the tendency for languages to order demonstratives (Dem), numerals (Num), and adjectives (Adj) in a particular fashion, favoring the following linear orders within the noun phrase (NP): Dem-Num-Adj-N, N-Dem-Num-Adj, or N-Adj-Num-Dem. While the typological literature since Greenberg (1963) has in fact attested a wide variety of NP-internal word orders, the Dem-Num-Adj-N and N-Adj-Num-Dem orders are estimated to account for nearly half of the world's languages, with the latter being the most prevalent (Dryer, 2018).

Recent ALL work has attributed these two dominant patterns to a bias for *scope-isomorphism*, a preference for syntactic structures to mirror underlying semantic scope relations—in this case, that demonstratives take scope over numerals and numerals over adjectives (Culbertson & Adger, 2014; Martin et al., 2019, 2020). Importantly, the Dem-Num-Adj-N and N-Adj-Num-Dem orders are both scope-isomorphic. Using an extrapolation variant of ALL (a.k.a. the poverty-of-the-stimulus paradigm), these studies taught participants single-modifier NPs in an artificial language, then examined how the participants ordered *multiple* modifiers during the critical testing stage. As is key to the extrapolation paradigm, the linguistic stimuli taught to the participants were always ambiguous as to whether the artificial language adhered to isomorphic or non-isomorphic linearizations, since the participants only ever saw one modifier within a given NP (i.e., N-Adj, N-Num, and N-Dem).

Across the board, Culbertson and Adger (2014) and Martin et al. (2019, 2020) found a preference for scope-isomorphic word orders: When the participants had to produce novel, multi-modifier NPs, they tended to order the words in a way that reflected semantic scope. All three studies were conducted with English speakers, with Martin et al. (2019) additionally focusing on Thai speakers. It's important to note that Martin et al. (2020) corrected for methodological issues in the first two studies, which we revisit in the Discussion.

Methods

To test the viability of a fully iconic artificial language, we sought to reproduce the scope-isomorphism preferences Culbertson and Adger (2014) and Martin et al. (2019, 2020) observed with English speakers. In particular, we paired pictographic linguistic stimuli with the visual stimuli from Martin et al. (2020). Each scene from Martin et al. (2020) depicted a table with a girl standing behind it. The participants were tasked with describing objects that appeared on the table. For example, if there were two feathers spread apart on the table and the girl was pointing to the closest one, this was meant to solicit the interpretation "this feather" (ordered N-Dem in the artificial language).

Adopting the extrapolation paradigm, we trained the participants on bare and one-modifier NPs using the

Table 1: The Pictographic Lexicon.

Noun	Adj	Num	Dem
ball	*red*	*two*	*this*
feather	*black*	*three*	*that*
mug			

aforementioned stimuli, then examined the participants' ordering preferences when tasked with producing two-modifier NPs. We taught all three modifier types—Dem, Num, and Adj—to each participant (akin to Experiment 2a in Culbertson & Adger, 2014). In the two-modifier production trials, we measured whether or not the productions were scope-isomorphic. As in Culbertson and Adger (2014) and Martin et al. (2019, 2020), we hypothesized that participants would produce scope-isomorphic modifier orderings at a rate significantly greater than chance, and that this would be true for all types of two-modifier NPs.

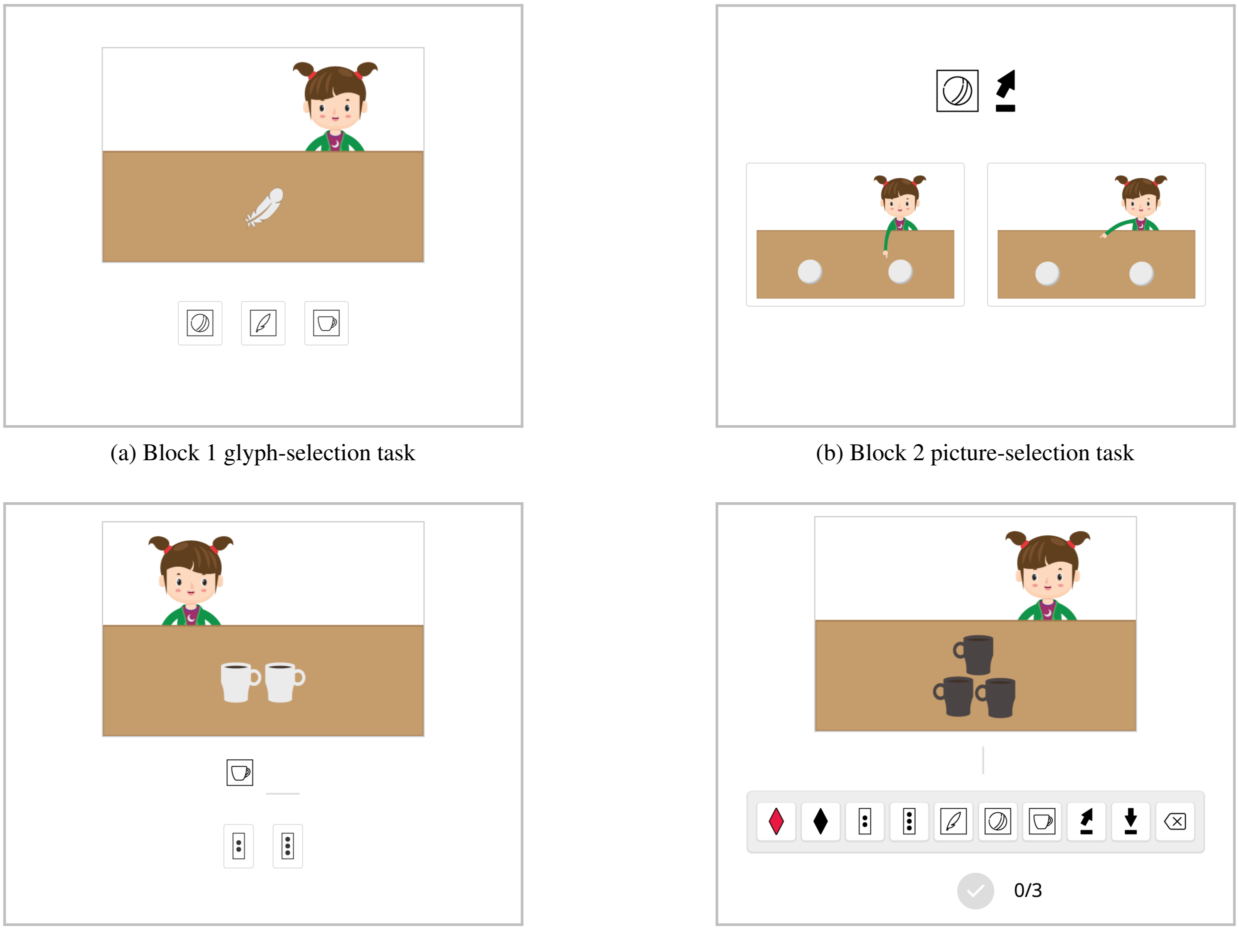
The Iconic Language

We created a pictographic lexicon to represent the 3 nouns and 6 modifiers (2 Dem, 2 Num, 2 Adj) from Martin et al. (2020), shown in Table 1. These stimuli were never phonologically realized in our study. The icons—hereafter, *glyphs*—were adapted from SVG icons downloaded from Flaticon (www.flaticon.com). The lexical items privileged within-category similarity. For instance, all of the noun glyphs were framed in squares and we used colored rhombuses for adjectives. Furthermore, we made all of the modifiers approximately the same width and height to circumvent visual biases that may lead participants to place modifiers closer to or farther away from the noun based on size. Differentiating the artificial language from English, the participants were taught noun-initial word orders (i.e., N-Dem, N-Num, and N-Adj) to avoid transfer effects, following earlier studies.

Procedure

We constructed the experiment in jsPsych (de Leeuw, 2015) using custom plugins. The experiment was composed of forced-choice and production-style exercises. There were 3 training blocks (26 trials) and 1 testing block (24 trials), totaling 4 blocks altogether (50 trials). Code for the experiment is available on GitHub (https://github.com/tsnaomi/iconic-all).

Instructions At the start of the experiment, the participants were told they would be learning a "pretend pictographic language". The instructions featured a mild deception that led the participants to believe they would be testing a new language learning app. This deception was done to help the participants buy into the language-learning exercise without overly analyzing the language itself; we disclosed the deception at the end of the study. In the instructions, we



(a) Block 1 glyph-selection task

(b) Block 2 picture-selection task

(c) Block 2 cloze task

(d) Block 4 production task

Figure 1: Experiment trials with visual stimuli from Martin et al. (2020).

explicitly defined the Dem glyphs as *this* and *that* to encourage determiner interpretations of these items.[2] However, we refrained from giving English translations for the other vocabulary to minimize the participants' English awareness during the experiment.

Since one of our goals was to make the study as language-neutral as possible, only the consent form and the instructions at the start of the study, as well as a post-experiment questionnaire, were presented in English (i.e., language-specific). No other English text appeared during the study, with the arguable exception of Arabic numerals in Blocks 3 and 4. This had the added benefit of minimizing English priming and processing during the experiment. In lieu of providing English instructions during the experiment trials, we used pictures and simple CSS animations to guide the participants through the study. For instance, in each forced-choice trial, the different options would be surrounded by a glow until an item was selected.

[2]While piloting the experiment, we found that participants favored directional/prepositional interpretations such as *down* and *across* when they were not given the Dem meanings in advance.

Training Blocks We implemented an "active learning" design, where we refrained from *teaching* the participants the meaning of the lexical items upfront (e.g., we did not tell them ♦ means *red*). Instead, the training trials immediately quizzed the participants on the artificial language, requiring the participants to intuit the meanings of the glyphs—which was made easy by their iconicity. At the end of each trial, the participants were informed as to whether they had answered correctly and were further required to correct any mistakes. For example, when an incorrect answer was selected in the forced-choice exercises, it would become outlined in red. The correct answer would then glow green, bouncing up and down in a "pick me" animation until the participant clicked on it, advancing them to the next trial.

Block 1 of the experiment focused on noun learning. For each of the 3 nouns, the participants completed a glyph-selection task and a picture-selection task, totaling 6 trials. In the glyph-selection tasks, the participants were shown a picture of an object, then had to select the corresponding glyph from the set of three noun glyphs (Figure 1a). Conversely, in the picture-selection tasks, the participants were shown a noun glyph and had to select the corresponding picture from the set of three noun images. The trials were intermixed with respect to the noun and task type, and the order of the selection choices were shuffled from trial to trial.

Block 2 introduced the 6 modifiers. For each modifier, the participants completed a picture-selection task and a cloze ("fill-in-the-blank") task, totaling 12 trials. In the picture-selection tasks (Figure 1b), the participants were shown a complete one-modifier NP and had to select the corresponding picture from a set of two pictures; the foil image was always the same noun paired with the other modifier of the same category. In the cloze task (Figure 1c), the participants were shown a picture with an incomplete one-modifier NP caption, then had to select the missing modifier given the choices of the correct modifier and the other modifier of the same category. Note that both trial types further served to familiarize the participants with the noun-initial word order. Each noun appeared four times across the block. The trials were intermixed with respect to the noun, modifier, and task type, and the order of the selection choices were shuffled.

In Block 3, the final training block, the participants practiced producing zero- and one-modifier NPs through 8 "keyboard" trials. In addition to reinforcing the noun-initial word order, this block familiarized the participants with the production task format. In each trial, the participants were presented with a picture, then tasked with producing a caption for the image using a clickable keyboard provided on the screen (cf. Figure 1d). The keyboard contained all nine glyphs in the lexicon, shuffled within category *and* across categories from trial to trial; a "backspace" key appeared on the far right. The participants were only able to submit a response once they had entered the correct number of glyphs, as indicated by a glyph counter at the bottom of the screen (presented in Arabic numerals). If a participant entered the wrong caption, they were shown the correct caption and prompted to correct their answer, which then allowed them to proceed to the next trial. The block began with 2 randomly-selected bare noun trials, followed by 6 one-modifier NP trials. Note that, given the lexicon's 3 nouns and 6 modifiers, there were 18 possible one-modifier NPs; accordingly, the 6 one-modifier NPs presented in Block 3 were the ones held out from Block 2.

Testing Block Block 4 tested the participants on one- and two-modifier NP productions using the same keyboard task design from Block 3 (Figure 1d). The block consisted of 12 *non-critical* one-modifier NP trials and 12 *critical* two-modifier NP trials, totaling 24 trials. The block began with 4 randomly-selected non-critical trials, followed by the remaining 8 non-critical trials and the 12 critical trials intermixed. The 12 one-modifier NPs were the ones observed in Block 2, meaning that, across the experiment, the participants *produced* each one-modifier NP exactly once. The 12 two-modifier NPs included each possible combination of the 6 modifiers, with each noun appearing 4 times. As in Block 3, the participants could only submit a response once they had entered the correct number of glyphs. This was done to ensure that the participants provided *maximal* descriptions of the images. In contrast to Block 3, the participants were not given feedback on their responses.

Table 2: Logistic Regression Results.

(Intercept)	$\hat{\beta}$	*SE*	*p*
All data	2.27	0.36	<0.001
Dem-Adj	7.78	1.70	<0.001
Dem-Num	9.92	2.08	<0.001
Num-Adj	2.30	0.96	0.017

Counterbalancing In Blocks 3 and 4, the keyboards were counterbalanced such that the relevant keys appeared in matching order of the expected answer in half the trials, and in the reverse order in the remaining trials. In the one-modifier NP trials, this meant that the noun appeared before the correct modifier in half of the trials and after the modifier in the other half. Likewise, in the two-modifier NP trials, the modifiers in question appeared in isomorphic order in half of the keyboards and in non-isomorphic order in the other half.

Post-Experiment Questionnaire After the experiment, the participants completed a brief questionnaire, which asked them to give English translations for each glyph and inquired after the strategies they used during the study, particularly with respect to how they ordered the modifiers during the testing block. We further asked the participants about the extent to which they "verbalized" the icons to themselves and in what languages. The questionnaire concluded with a language history form loosely inspired by the LEAP-Q (Marian, Blumenfeld, & Kaushanskaya, 2007).

Participants

We recruited 55 participants online via the Prolific platform (www.prolific.co). Using Prolific's pre-screening filters, we only recruited participants who self-identified as monolingual English speakers and who reported having no language-related disorders or issues seeing colors. All of the participants gave informed consent and received 4 USD in compensation.

Based on their Block 4 responses, we included 45 of the participants in our analysis. We required participants to produce at least 10 correct one-modifier NPs (out of 12; 83%) and at least 9 analyzable two-modifier NPs (out of 12; 75%). For the non-critical NPs, a response was marked as incorrect if it included the wrong glyphs or if the glyphs appeared in the wrong order. Likewise, for the critical NPs, a response was marked as un-analyzable if it contained the wrong glyphs or if the noun did not appear phrase-initially. The participants who qualified for the analysis took on average 7.0 minutes to complete the experiment (range: 3.3-33.4) and 4.8 minutes to complete the post-experiment questionnaire (range: 1.2-24.8).

Results

Out of the analyzable two-modifier NPs, 82.3% were in scope-isomorphic order (Dem-Adj: 87.3%, Dem-Num: 90.6%; Num-

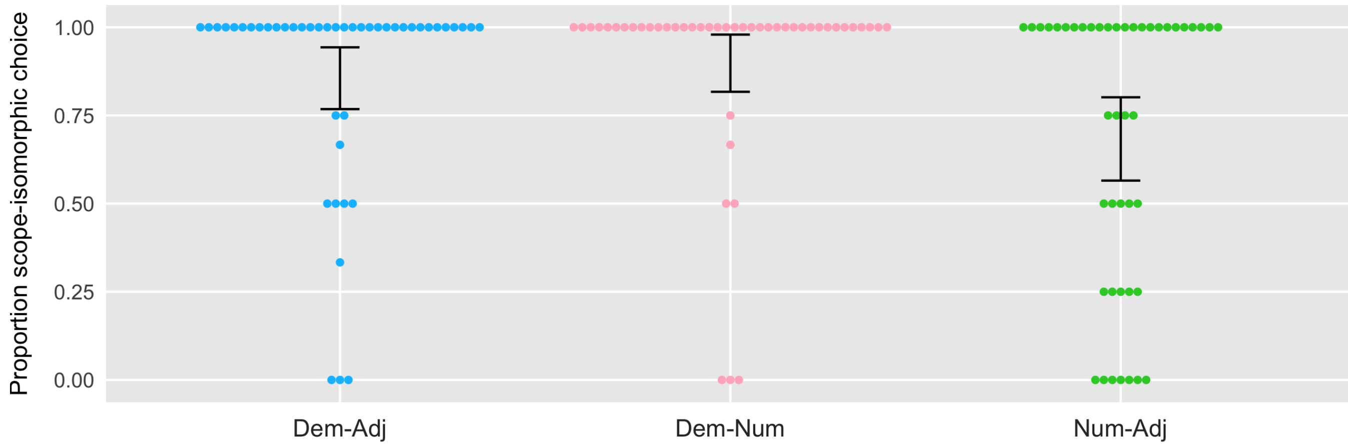


Figure 2: Proportion of responses that were scope-isomorphic (with 95% confidence intervals). Each dot represents a participant.

Adj: 69.3%). *By participant*, the average percentage of the responses that were scope-isomorphic was 81.2% (Dem-Adj: 85.6%; Dem-Num: 89.8%; Num-Adj: 68.3%), visualized by modifier group in Figure 2.

We fit a logistic mixed-effects regression model to the critical NPs, predicting a response's modifier order (1 = *isomorphic*; 0 = *non-isomorphic*). We included random intercepts for participants, but did not do so for each noun, since there was zero variance. We found that the intercept was positive and significant (top row of Table 2), indicating that the participants chose scope-isomorphic modifier orders at above chance level.

In a likelihood ratio test, we further compared the model to one that included a fixed effect for modifier group (Dem-Adj, Dem-Num, and Num-Adj) and found that the latter model better fit the data ($p < 0.001$), indicating that the preference for scope-isomorphic orders was stronger for some modifier pairs than others. To facilitate interpretation, we then fit separate models for each group, confirming that the isomorphism preference exists within each modifier pair (bottom three rows of Table 2).

Discussion

We set out to test the viability of iconic stimuli in ALL. We attempted to use a pictographic lexicon to replicate prior findings showing that speakers favor modifier orders that preserve semantic scope relations (Culbertson & Adger, 2014; Martin et al., 2019, 2020). Consistent with this work, our experiment showed that participants were significantly more likely to produce scope-isomorphic orders, validating the utility of iconic artificial languages.

Semantic Scope

The present study further supports a preference for scope-isomorphic word orders within the noun phrase, at least among English speakers. Furthermore, we found that the strength of this bias was modulated by modifier pair, with comparable intercepts observed among the Dem-Adj and Dem-Num groups, and a smaller intercept in the Num-Adj group.[3]

On the surface, these findings differ modestly from those of Culbertson and Adger (2014) and Martin et al. (2020): When running separate *conditions* for each modifier pair, both sets of authors saw significantly higher proportions of scope-isomorphic orders in the Dem-Adj condition, with the Dem-Num condition instead patterning more similarly to the Num-Adj condition. In response to these findings, Culbertson and Adger (2014) hypothesized that a preference for scope-isomorphism for a subset of modifiers (e.g., Dem-Adj) may lead to a stronger scope-conforming mapping for *all* of the modifiers. In follow-up experiments, they taught all three modifiers to the same set of participants—akin to the present study. When they did so, the proportion of scope-isomorphic responses no longer differed significantly between the groups.

One possibility is that teaching all three modifier categories to our participants similarly resulted in a stronger scope-conforming preference in the Dem-Num group. At the same, since our experiment was much shorter than Culbertson and Adger's, it may be that our participants were not "immersed" enough in the pictographic language to develop an equally strong scope-isomorphism preference in the Adj-Num group.

Interestingly, Martin et al. (2020) suggested that the observed between-condition effects may reflect differences in the "strength of associations" held between the lexical categories. Citing Culbertson, Schouwstra, and Kirby (2020), who quantified these associations with pointwise-mutual information, Martin et al. speculated that "this kind of dependency determines how likely speakers are to separate a head and dependent". Notably, Culbertson et al. (2020) had found that N$\sim$Adj, on average, were most closely associated, followed by N$\sim$Num and then N$\sim$Dem. Building on Martin et al.'s proposal, it's possible that when taught all three modifier categories, the amount of exposure required to form

[3] A post hoc analysis that excluded Num-Adj data found no significant difference between the Dem-Adj and Dem-Num groups.

a strong scope-conforming mapping between two particular modifiers is relative to the same strength of associations. We leave this question for future work. However, if true, the present findings would further demonstrate the ability of iconic artificial languages to test fine-grained linguistic hypotheses.

Finally, a remaining question is whether the bias for scope-isomorphism is universal or instead reflects structural transfer from the participants' native languages, since English NPs are scope-isomorphic. Confirmed in personal communication, Martin and colleagues are currently addressing this question with speakers of Kîîtharaka, a Bantu language where the surface word order is non-scope-conforming (though see Kanampiu & Muriungi, 2019). If Kîîtharaka speakers do not display a preference for scope-isomorphic orders, this could suggest that the bias is *not* universal and that the present findings results from crosslinguistic transfer; alternatively, it could be that the bias is universal but frequent experience with non-scope-isomorphic linearizations can dampen it. On the other hand, if Kîîtharaka speakers do display a preference for scope-isomorphic orders, it would be interesting to see whether the *strength* of this preference is attenuated by the participants' knowledge of Kîîtharaka—in other words, to see if there is still an effect of crosslinguistic influence. Importantly, by allowing the same artificial language to be used with multiple language populations, the iconic ALL paradigm proposed here provides a promising vehicle for measuring and juxtaposing such effects of crosslinguistic influence.

Future Directions

Perhaps ironically, the absence of a phonologically-realized lexicon may inadvertently lead to transfer effects if participants are *verbalizing* the iconic stimuli to themselves in their primary languages. Indeed, our post-experiment questionnaire revealed that the participants often 'said the symbols out loud' or 'in their head', as summarized in Figure 3. This could have resulted in unwanted transfer effects from English.

However, it's worth highlighting that the experiments in Culbertson and Adger (2014) and Martin et al. (2019) involved similar noun-initial NPs, but used English words for English speakers and Thai words for Thai speakers (e.g., teaching participants *ball that* to mean "that ball"). As discussed in Martin et al. (2020), this led many participants to adopt a "flipping" strategy, where they arrived at scope-isomorphic NPs simply by reversing English word order. Martin et al. therefore used an artificial lexicon with nonce modifiers (e.g., *puku*), finding still a scope-isomorphic preference, but without the confound of the participants transferring their English knowledge to the task.

Similarly, only one participant in the present study reported using a flipping strategy, with the majority of the participants reporting that they simply picked a word order and stuck with it. This shows that, despite the participants articulating the iconic lexical items to themselves, this did not trigger the same level of crosslinguistic influence encountered by Culbertson and Adger (2014) and Martin et al. (2019). Yet,

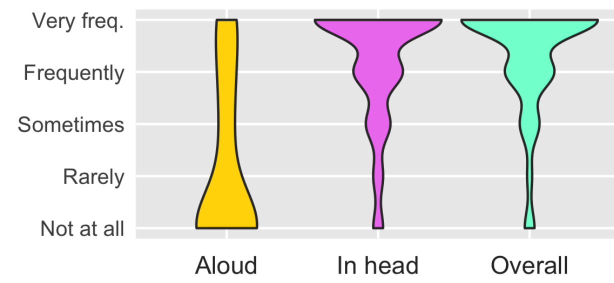


Figure 3: How often the participants "verbalized" the glyphs. *Overall* takes the by-participant max for *Aloud* and *In head*.

if verbalizing iconic stimuli *can* elicit transfer effects, if exploited strategically, this provides an avenue for studying and comparing crosslinguistic influence across different languages, while holding the artificial language constant.

Still, it is important to point out that a truly neutral artificial language is likely impossible. Depending on the goal of the study, an iconic artificial language may need to be adjusted from language to language, particularly at the level of word order. For instance, if we were to replicate the scope-isomorphism bias with Thai speakers, we would want to reverse the word order taught in the present experiment, training the participants on noun-*final* NPs, since Thai NPs are noun-*initial* (cf. Martin et al., 2019). Furthermore, *language-neutral* does not mean *culture-neutral*, since pictographic stimuli can still involve culturally-grounded semiotics. Nevertheless, a thoughtfully designed pictographic lexicon can enable more finely controlled comparisons across different languages.

To summarize, future ALL work should draw on iconicity to study more typologically diverse language populations. Given the relative language-neutrality of pictographic stimuli, the same artificial lexicon can be used with speaking and signing populations, typologically and orthographically diverse languages, children and adults, as well as monolinguals, bilinguals, and multilingual individuals. Such studies should delve into how these communities both *resemble* and *vary* from one another in their language learning and processing.

Conclusion

Endeavoring towards language-neutral stimuli in artificial language learning, we have shown the viability of replacing phonologically-realized writing systems with fully iconic, non-orthographic symbols. Performing a successful conceptual replication of experiments by Culbertson and Adger (2014) and Martin et al. (2019, 2020), we verified that English speakers prefer modifier orders that comport with semantic scope relations ("scope-isomorphism")—even when learning noun-initial NPs composed entirely of icons. Crucially, iconic artificial languages make it possible to re-use the same linguistic stimuli with diverse language populations, facilitating crosslinguistic investigations. We encourage future research to do just that.

References

Bowerman, J., & Smith, K. (2022). An experimental study of semantic extension in a novel communication system. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *44*(44).

Culbertson, J. (2023). Artificial language learning. In J. Sprouse (Ed.), *Oxford Handbook of Experimental Syntax* (pp. 271–301). Oxford University Press.

Culbertson, J., & Adger, D. (2014). Language learners privilege structured meaning over surface frequency. *Proceedings of the National Academy of Sciences*, *111*(16), 5842–5847.

Culbertson, J., & Kirby, S. (2022). Syntactic harmony arises from a domain-general learning bias. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *44*(44).

Culbertson, J., Schouwstra, M., & Kirby, S. (2020). From the world to word order: deriving biases in noun phrase order from statistical properties of the world. *Language*, *96*(3), 696–717.

Dautriche, I., Buccola, B., Berthet, M., Fagot, J., & Chemla, E. (2022). Evidence for compositionality in baboons (Papio papio) through the test case of negation. *Scientific Reports*, *12*(1), 19181.

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a web browser. *Behavioral Research Methods*, *47*(1), 1–12.

Dryer, M. S. (2018). On the order of demonstrative, numeral, adjective, and noun. *Language*, *94*(4), 798–833.

Fedzechkina, M., Newport, E. L., & Florian Jaeger, T. (2016). Miniature artificial language learning as a complement to typological data. In L. Ortega, A. E. Tyler, H. I. Park, & M. Uno (Eds.), *The usage-based study of language learning and multilingualism* (pp. 211–232). Georgetown University Press.

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.

Greenberg, J. H. (1963). *Universals of language*. Cambridge, Massachusetts: MIT Press.

Kanampiu, P. N., & Muriungi, P. K. (2019). Order of modifiers in Kiitharaka determiner phrase. *International Journal on Studies in English Language and Literature*, *7*, 10-21.

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research*.

Martin, A., & Culbertson, J. (2020). Revisiting the suffixing peference: Native-language affixation patterns influence perception of sequences. *Psychological Science*, *31*(9), 1107–1116.

Martin, A., Holtz, A., Abels, K., Adger, D., & Culbertson, J. (2020). Experimental evidence for the influence of structure and meaning on linear order in the noun phrase. *Glossa*, *5*(1), 1–21.

Martin, A., Ratitamkul, T., Abels, K., Adger, D., & Culbertson, J. (2019). Cross-linguistic evidence for cognitive universals in the noun phrase. *Linguistics Vanguard*, *5*(1), 20180072.

Perniss, P., Thompson, R. L., & Vigliocco, G. (2010). Iconicity as a general property of language: Evidence from spoken and signed languages. *Frontiers in Psychology*, *1*.

Saratsli, D., Bartell, S., & Papafragou, A. (2020). Cross-linguistic frequency and the learnability of semantics: Artificial language learning studies of evidentiality. *Cognition*, *197*, 104194.

Sato, A., Schouwstra, M., Flaherty, M., & Kirby, S. (2020). Do all aspects of learning benefit from iconicity? Evidence from motion capture. *Language and Cognition*, *12*(1), 36–55.