Who benefits from redundancy in learning noun class systems?

Aislinn Keogh*[1] and Gary Lupyan[2]

*Corresponding Author: aislinn.keogh@ed.ac.uk
[1]Centre for Language Evolution, University of Edinburgh, Edinburgh, UK
[2]Department of Psychology, University of Wisconsin-Madison, Madison, WI, USA

Redundancy is ubiquitous in the world's languages, but its functions are not yet well understood. Here, we propose that redundancy might contribute to the robustness of language to facilitate its learning by users with diverse cognitive traits. We use an artificial language learning experiment to identify individual differences in learning of noun classes from redundant linguistic cues. All logically possible behaviors are represented in our data: some participants prefer Cue A, some prefer Cue B, and some form a more holistic representation of Cue A+B. Despite this diversity, the population as a whole was above-chance when generalizing to novel stimuli, suggesting that redundancy helps people converge on similar surface structures, even if their underlying representations differ.

1. Introduction

All languages have a substantial degree of redundancy: the same information is encoded in multiple parts of the signal. For example, morphosyntactic elements such as agreement systems often involve marking words for features (e.g. person, gender, number or case) that are predictable from other cues (Haig & Forker, 2018). The pervasiveness of redundancy in language is a puzzle, especially in the face of evidence that producers prefer to minimize redundancy by omitting or reducing more predictable elements (e.g. Gibson et al., 2019; Jaeger, 2010; Aylett & Turk, 2004). One proposed explanation is that redundancy is a design feature that improves language learning, especially for young children (Tal & Arnon, 2022; Lupyan & Dale, 2010; Gerken et al., 2005; Morgan et al., 1987; Portelance et al., 2023) or the real-time processing of language (Christiansen & Chater, 2016). Another (non-mutually exclusive) possibility is that redundancy contributes to the *robustness* of language in the face of having to be acquired by diverse learners (Monaghan, 2017; Winter, 2014; Whitacre, 2010). Although there are cultural selection pressures against language structures that are not learnable by a large proportion of the population (Kirby et al., 2015), language systems may not be able to be optimized to be equally learnable by *all* members of a community, given the diversity in people's cognitive traits. Redundancy may be one way to increase the likelihood that a language will be learned equally well by everyone: even if some people fail to learn certain cues, they should still be able to learn the

linguistic system overall.

Here, we offer an exploratory analysis of individual differences in learning of a redundant system. We present an experiment in which participants are trained on an artificial language with noun classes marked by redundant linguistic cues: a suffix on the noun, and a separate class marker. We then test how well they have learned these cues by asking them to generalize to novel meanings. Naturally, we expect to see variability in how well people learn the training set. However, our key interest is whether there is also variability in *generalization*, even among participants who appear to be performing similarly in training. A range of behaviors are logically possible: participants could learn the training set by rote without identifying any underlying rules or structure, they could learn both cues to class membership equally well, or they could learn the two cues to differing extents. We consider participants' training profiles and cognitive dispositions to try and predict who is more likely to exhibit these different behaviors.

2. Method

Participants We recruited 100 adults via Prolific. All resided in the US and were self-reported native English speakers with no known language disorders. Participants were paid $7.30 for around 45 minutes' participation.

Materials Stimuli were drawings from the MultiPic databank (Duñabeitia et al., 2018). We selected eight basic-level categories from four semantic domains: humans, animals, food, and clothing. The lexicon consisted of 32 pseudoword roots, four suffixes and four class markers taken from Culbertson et al. (2017). A full phrase consisted of a pre-nominal class marker followed by root + suffix e.g. *gae skun-po*. Phrases were displayed both auditorily and orthographically.

Procedure The experiment was written in JavaScript using the jsPsych library (de Leeuw, 2015) and administered through participants' web browser. First, participants were trained on a subset of the artificial language: four randomly selected meanings from each class. On each trial, participants heard a phrase and attempted to select its meaning from a 2x2 array of images: two from the target class, and two from another randomly selected class. They received full feedback on their selection. Participants completed 8 blocks of training, with each of the 16 meanings appearing as the target on one trial per block (128 trials total). Next, participants completed a reading span task to provide a measure of verbal working memory (Daneman & Carpenter, 1980; Friedman & Miyake, 2005), and a questionnaire assessing approach and avoidant behavioral tendencies (the BIS/BAS measurement tool: Carver & White, 1994). Both of these variables have been found to correlate with generalization performance in other domains (e.g. Dale et al., 2021). Participants were then tested on their knowledge of the language's structure using the held-out meanings. On each trial, participants heard an unfamiliar phrase and

attempted to select its meaning from a 2x2 array of images: the target, and one randomly selected meaning from each of the other classes. There were three trial types in this phase. On REDUNDANT trials participants saw complete phrases as in training; on CLASSIFIER-ONLY and NOUN-ONLY trials they saw only one cue (the missing word was blanked out). They received no feedback on their selections. Finally, participants completed a questionnaire assessing explicit awareness of the noun classes and other language learning experience.

3. Results

Training Overall, participants showed clear evidence of learning over the course of training, with accuracy increasing considerably from the first block (M = 0.43, SD = 0.49) to the final block (M = 0.84, SD = 0.37). However, there are noticeable differences between participants. Visual inspection of by-participant loess curves reveals that there are at least three qualitatively different training profiles (Fig. 1A): linear (accuracy continues to increase throughout the training phase), logistic (accuracy increases from the start of training but ultimately reaches an asymptote) and non-monotonic (accuracy varies across the training phase). Controlling for explicit awareness of the semantic categories, higher performance in the final training block was predicted by higher performance in the first block (β = 0.041, SE = 0.015, t = 2.72, $p < 0.01$), higher working memory capacity (β = 0.036, SE = 0.016, t = 2.27, $p < 0.05$) and greater experience with language-learning apps like Duolingo (β = 0.032, SE = 0.015, t = 2.14, $p < 0.05$).

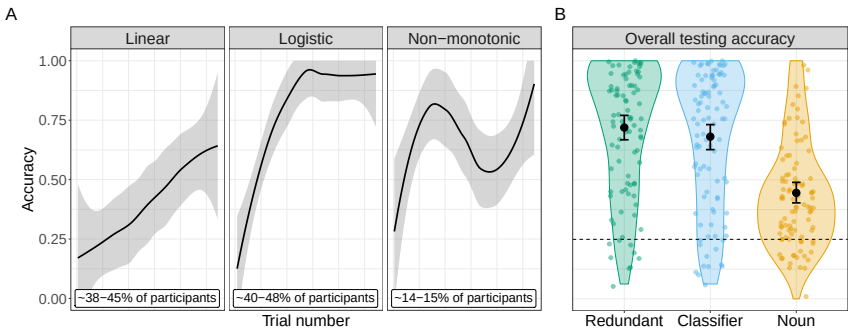

Figure 1. **A:** Example training profiles from three characteristic participants. **B:** Accuracy on test trials by cue. Individual coloured points represent by-participant mean performance for that cue. Black points and error bars represent the mean and bootstrapped 95% confidence interval over participants. The dashed line indicates chance performance of 0.25.

Generalization Because participants had not previously been exposed to the specific meanings presented at test, they could not rely on knowledge of the roots to determine the meaning of the phrases. Rather, the only way they could succeed at this task was if they had learned the relationship between the classifier and/or suffix and the semantic class. A larger drop in accuracy between the final training block and the test phase indicates that a person is less good at learning that relationship. Importantly, someone can be very good at memorizing the specific phrase-item pairings presented in training (high training accuracy) and yet fail to generalize, resulting in low accuracy on the test trials.

Overall, performance was above chance for all trial types, indicating that at the population-level, there is generalization of class cues (Fig. 1B). Accuracy was highest for REDUNDANT trials (M = 0.72, SD = 0.45), closely followed by CLASSIFIER-ONLY trials (M = 0.68, SD = 0.47). NOUN-ONLY trials had considerably lower accuracy (M = 0.45, SD = 0.50). Unsurprisingly, accuracy on the final training block predicted overall test performance (t = 3.47, $p < 0.001$). Surprisingly, this relationship (r = 0.11) all but disappeared when we include measures of working memory (β = 0.075, SE = 0.021, t = 3.530, $p < 0.001$), risk aversion (β = 0.043, SE = 0.019, t = 2.31, $p < 0.05$) and explicit awareness of the association between word forms and semantic categories (β = 0.099, SE = 0.040, t = 2.45, $p <$ 0.05). These three covariates all independently predicted test performance while controlling for final training block accuracy (β = 0.036, SE = 0.031, t = 1.796, p = 0.076), together accounting for 32% of the variance. Thus, better learners were not necessarily better generalizers.

Unsurprisingly, there was a clear drop-off in accuracy from the final training block to test (averaging across trial types: M = -0.22, SD = 0.23). A larger drop (controlling for training performance) is consistent with people being more focused on memorizing the specific items than on learning the underlying rules. Looking just at REDUNDANT test trials (the most like-for-like comparison), higher working memory capacity was associated with a smaller drop-off (β = -0.11, SE = 0.025, t = -4.23, $p < 0.001$). Higher reward responsiveness was associated with a slightly *larger* drop-off (β = 0.051, SE = 0.023, t = 2.21, $p < 0.05$), potentially due to the lack of feedback during testing (positive feedback may be viewed as a kind of reward).

Almost no one had uniform performance on the three test trial types, suggesting that the vast majority of participants had a preferred cue. We calculated two indices for every participant to compare their average performance on REDUNDANT trials to each of the individual cues. A larger positive score for the comparison between Cue A and REDUNDANT trials indicates that a participant is relying more on *Cue B*, since their performance is more greatly impaired by the removal of that cue. A variety of behaviors are represented in our data (Fig. 2), but participants clearly tend to rely more on the classifier than the suffix. Only 21 participants performed equally well or better on NOUN-ONLY trials relative

to REDUNDANT trials, compared to 55 who performed equally well or better on CLASSIFIER-ONLY trials. Many (32) participants had positive scores on both indices, indicating that they were specifically benefiting from the redundancy i.e. had learned the association between classifier+suffix and semantic category in a more holistic way such that their performance declined when either cue was missing. Controlling for raw performance on REDUNDANT trials, higher performance in the final training block was associated with an increased benefit of redundancy (β = 0.047, SE = 0.011, t = 2.96, $p < 0.01$), while higher performance in the first half of the training phase (i.e. faster learning) was associated with a reduced benefit of redundancy (β = -0.043, SE = 0.012, t = -3.15, $p < 0.001$). In fact, faster learners actually performed *worse* when presented with redundant cues compared to their preferred cue in isolation. Greater risk aversion was also associated with a lower redundancy advantage (β = -0.028, SE = 0.010, t = -2.88, $p < 0.01$).

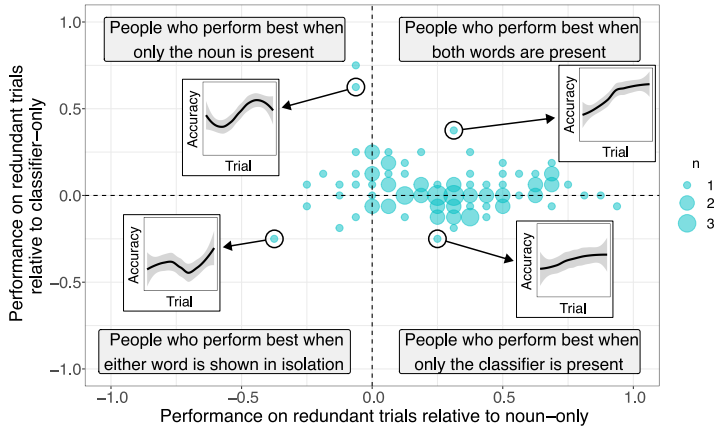

Figure 2. Performance on REDUNDANT trials relative to individual cues. Individual points represent by-participant scores; larger points represent more participants with equivalent values. Positive scores indicate a facilitatory effect of redundancy; negative scores indicate a detrimental effect of redundancy. Points along the dashed lines indicate that performance is equally good on REDUNDANT trials as on the given individual cue. Insets show the learning curves for the highlighted participants.

4. Discussion

In this study, we investigated whether morphosyntactic redundancy could contribute to the robustness of language by providing greater assurance that a system will be acquired despite variability in learning mechanisms across a population. When trained on an artificial language with two linguistic cues to noun class membership, we found clear individual differences in cue preference. Although the

majority of our participants relied more heavily on the separate class-marker, a sizeable minority were attending more to the suffix on the noun. Around a third of our participants showed evidence of integrating the two cues more closely, performing best when both cues were available. This redundancy benefit was greatest for participants who achieved the highest level of accuracy by the end of the training phase, and lowest for participants who reached a higher level of accuracy earlier on in training.

The lack of a redundancy advantage for faster learners suggests that early commitment to one cue that reliably predicts category membership may block discovery of additional generalizations that might be beneficial down the line (in classical conditioning terms, *overshadowing*: Pavlov, 1927). Learners who explore the data for longer may be better able to integrate the redundant cues, and use these extra sources of information to their advantage both in learning and in generalization (Liquin & Gopnik, 2022; Sumner et al., 2019). Higher risk aversion also appears to reduce the strength of this overshadowing effect, resulting in more even performance across the three trial types, and therefore a lower benefit of redundancy *per se*. It is important to note that this is not a straightforward consequence of these participants expending greater effort: a person could be trying very hard during training, yet fail to learn the structure in a way that enables them to generalize training data effectively.

Contrary to some previous work in the 'Less-is-More' tradition (e.g. Goldowsky & Newport, 1993; Kareev, 1995; Pitts Cochran et al., 1999), we also found a positive relationship between working memory capacity and generalization. This finding dovetails with more recent work arguing that enhanced cognitive capacity is associated with better L1 and L2 learning outcomes (e.g. Brooks & Kempe, 2019; Rohde & Plaut, 2003), as well as studies linking higher working memory capacity to better category learning (e.g. Craig & Lewandowsky, 2012).

Our study also offers preliminary evidence of robustness effects in morphosyntax. Future work can implement different training conditions to see whether, at a population-level, there is better generalization of a language with redundant cues than one with a single cue – even if that single cue is well-learned by the majority of participants, like the classifier in our experiment. Manipulating the reliability of the redundant cues (Monaghan et al., 2017) may also force people to attend to both cues, reducing individual differences in cue preference. In fact, it is possible that even those participants who seemed to benefit from redundancy did not interpret the cues as redundant *per se*: since both were always available in training, they could have been interpreted as a single discontinuous cue.

Overall, this study adds to a growing body of evidence suggesting that, despite its potential costs in production, redundancy may be functional for language learning. Specifically, we suggest that when multiple cues to a language's grammatical structure are available, learners who favour different cues should nonetheless be able to acquire that underlying structure equally well.

Acknowledgements

This project received funding from the Economic and Social Research Council (ref. ES/P000681/1, held by AK) and NSF-PAC (ref. 2020969, held by GL). We are grateful to Kira Breeden for recording the audio stimuli.

References

Aylett, M., & Turk, A. (2004). The Smooth Signal Redundancy Hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech*, *47*(1), 31–56.

Brooks, P. J., & Kempe, V. (2019). More is more in language learning: Reconsidering the Less-Is-More Hypothesis. *Language Learning*, *69*, 13–41.

Carver, C. S., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS Scales. *Journal of Personality and Social Psychology*, *67*(2), 319–333.

Christiansen, M. H., & Chater, N. (2016). The Now-or-Never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences*, *39*.

Craig, S., & Lewandowsky, S. (2012). Whichever way you choose to categorize, working memory helps you learn. *Quarterly Journal of Experimental Psychology*, *65*(3), 439-464.

Culbertson, J., Gagliardi, A., & Smith, K. (2017). Competition between phonological and semantic cues in noun class learning. *Journal of Memory and Language*, *92*, 343–358.

Dale, G., Cochrane, A., & Green, C. S. (2021). Individual difference predictors of learning and generalization in perceptual learning. *Attention, Perception, & Psychophysics*, *83*(5), 2241–2255.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*(4), 450–466.

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

Duñabeitia, J. A., Crepaldi, D., Meyer, A. S., New, B., Pliatsikas, C., Smolka, E., & Brysbaert, M. (2018). MultiPic: A standardized set of 750 drawings with norms for six European languages. *Quarterly Journal of Experimental Psychology (2006)*, *71*(4), 808–816.

Friedman, N. P., & Miyake, A. (2005). Comparison of four scoring methods for the reading span test. *Behavior Research Methods*, *37*(4), 581–590.

Gerken, L., Wilson, R., & Lewis, W. (2005). Infants can use distributional cues to form syntactic categories. *Journal of Child Language*, *32*(2), 249–268.

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.

Goldowsky, B. N., & Newport, E. L. (1993). Modeling the effects of processing limitations on the acquisition of morphology: The less is more hypothesis. *The proceedings of the 24th Annual Child Language Research Forum*(February), 124–138.

Haig, G., & Forker, D. (2018). Agreement in grammar and discourse: A research overview. *Linguistics*, *56*(4), 715–734.

Jaeger, T. F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology*, *61*(1), 23–62.

Kareev, Y. (1995). Through a narrow window: working memory capacity and the detection of covariation. *Cognition*, *56*(3), 263–269.

Kirby, S., Tamariz, M., Cornish, H., & Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition*, *141*, 87–102.

Liquin, E. G., & Gopnik, A. (2022). Children are more exploratory and learn more than adults in an approach-avoid task. *Cognition*, *218*, 104940.

Lupyan, G., & Dale, R. (2010). Language structure is partly determined by social structure. *PLOS ONE*, *5*(1), e8559.

Monaghan, P. (2017). Canalization of language structure from environmental constraints: A computational model of word learning from multiple cues. *Topics in Cognitive Science*, *9*(1), 21–34.

Monaghan, P., Brand, J., Frost, R., & Taylor, G. (2017). Multiple variable cues in the environment promote accurate and robust word learning. In *Proceedings of the 39th Cognitive Science Society Conference.*

Morgan, J. L., Meier, R. P., & Newport, E. L. (1987). Structural packaging in the input to language learning: Contributions of prosodic and morphological marking of phrases to the acquisition of language. *Cognitive Psychology*, *19*(4), 498-550.

Pavlov, P. I. (1927). *Conditioned reflexes: an investigation of the physiological activity of the cerebral cortex.* London: Oxford University Press.

Pitts Cochran, B., McDonald, J. L., & Parault, S. J. (1999). Too smart for their own good: The disadvantage of a superior processing capacity for adult language learners. *Journal of Memory and Language*, *41*(1), 30–58.

Portelance, E., Duan, Y., Frank, M. C., & Lupyan, G. (2023). Predicting age of acquisition for children's early vocabulary in five languages using language model surprisal. *Cognitive Science*, *47*(9), e13334.

Rohde, D. L. T., & Plaut, D. C. (2003). Less is less in language acquisition. In P. Quinlin (Ed.), *Connectionist Modelling of Cognitive Development.* Hove, UK: Psychology Press.

Sumner, E., Li, A., Perfors, A., Hayes, B., Navarro, D., & Sarnecka, B. (2019). The exploration advantage: Children's instinct to explore allows them to

find information that adults miss. *PsyArXiv Preprints*.

Tal, S., & Arnon, I. (2022). Redundancy can benefit learning: Evidence from word order and case marking. *Cognition*, *224*, 105055.

Whitacre, J. M. (2010). Degeneracy: a link between evolvability, robustness and complexity in biological systems. *Theoretical Biology and Medical Modelling*, *7*(1), 6.

Winter, B. (2014). Spoken language achieves robustness and evolvability by exploiting degeneracy and neutrality. *BioEssays*, *36*(10), 960–967.