EFFECTS OF FREQUENCY AND DISTRIBUTION ON LEARNING DEFAULTS

K.Pertsova (pertsova@unc.edu), B. Prickett (bprickett@umass.edu), and E.Chen (estherchen@unc.edu)

Introduction

Many patterns in language are best stated using defaults; hence the ubiquity of such theoretical devices in linguistics as the **Elsewhere Principle** (Panini; Kiparsky, 1973), **Blocking** (Aronoff, 1976), **the Subset Principle** (Halle 1997), among others.

How minority defaults are learned has been a subject of debate.

- Such patterns have been claimed to exist in languages, cf. German plural (Marcus et al., 1995), Arabic Broken Plural (McCarthy & Prince, 1990),
  - BUT The status of German –s as minority default is controversial (Zaretsky et al., 2016, Hahn & Nakisa (2000), McCurdy, 2019)
  - Likewise, with the Arabic plural (Boudelaa & Gaskell, 2000)
- Minority defaults have also been claimed to be problematic for the connectionist models (Pinker & Ulman, 2002.)
  - BUT Hare, Elman & Daugherty (1993) and Forrester & Plunkett (2019) claim to successfully model minority defaults using networks..
  - Recent work (McCurdy et.al. 2020) using neural Encoder-Decoder networks on German plural show that they do not generalize in the same manner as speakers

Broad research questions

1. How do humans learn default patterns and can they learn minority defaults?
2. Can current ED models learn minority defaults? Do they behave similarly to humans?

Why not look at German?
- there's no agreement on what the default is; and -there are many studies with inconclusive results, + non-deterministic pattern

METHODS:
1. Artificial grammar learning (AGL)
2. Modeling using neural nets

PREVIOUS FINDINGS WITH HUMANS
- Zaretsky et al. (2016) German children acquire high-frequency suffixes first
- McCurdy (2019) find that both –en (most frequent) and –s (widely distributed) suffix in German are extended to non-Rhymes
- Nevat et.al. (2018) using an AGL find that most frequent suffixes are overgeneralized at first, phon.distribution starts affecting generalization late in learning and is correlated with increased sensitivity to class-internal cues.

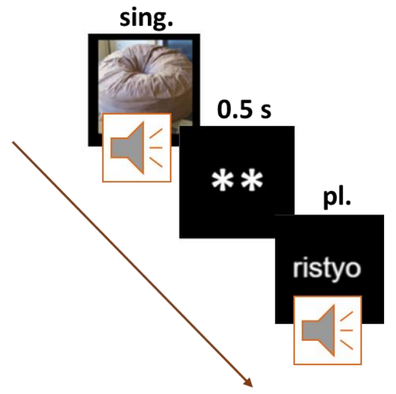
SPECIFIC QUESTIONS
- How is learning of defaults affected by
  - **Frequency:** equal frequency default vs. minority default
  - Learning strategy: **implicit** vs. **explicit**
    - H: defaults are learned faster and better by explicit learners

AGL Experiment

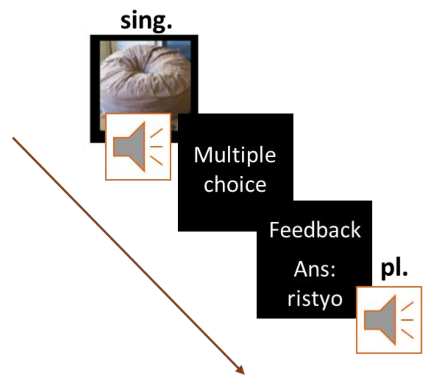
Stimuli: sg-pl pairs with plural allomorphy
- random assignment of 3 suffixes (-yo, -wa, and -ler) to 3 groups

Stimuli	Equal frequency	Minority default
Group 1, narrow 2 syll, ending in –N e.g, *ranom*	Freq: ~33%	Freq: high(45%) or mid (35%)
Group 2, narrow 1 syll, ending in -Ct e.g., *boft*	Freq: ~33%	Freq: high(45%) or mid (35%)
Group 3, wide Elsewhere e.g., *trofa*	Freq: ~33%	Freq: low (20%)

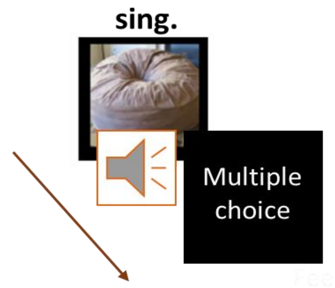
Training Phase 1



Training Phase 2



Testing Phase



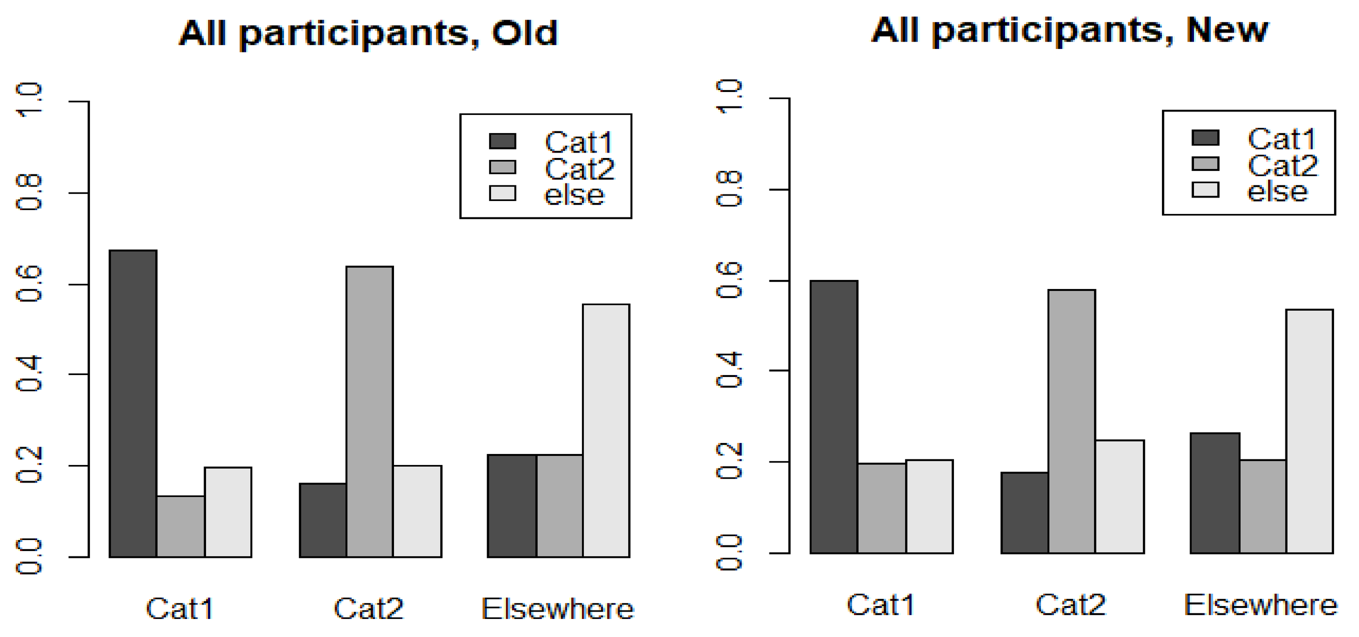
Post-experiment survey: participants stated what rules they found (if any)
- Rule-staters: those who at least stated partially-correct rule

Training Phase 1: 39-40 pairs
Training phase 2: same as phase 1
Testing phase: 18 trained words, 42 new words

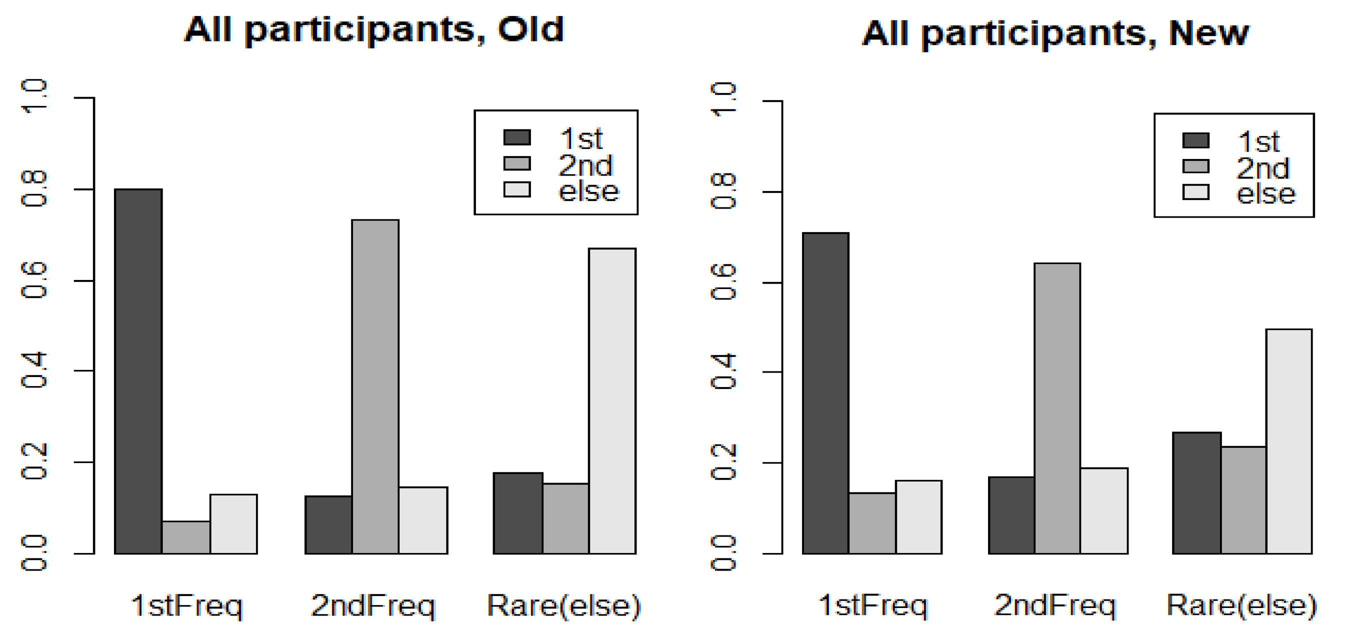
Do real languages have such patterns?

In English, comparative *–er* is restricted to monosyllabic adj or disyllabic adj ending in -[i]
- *blatanter, *beautifuler

Equal Frequency Condition



Minority Default Condition

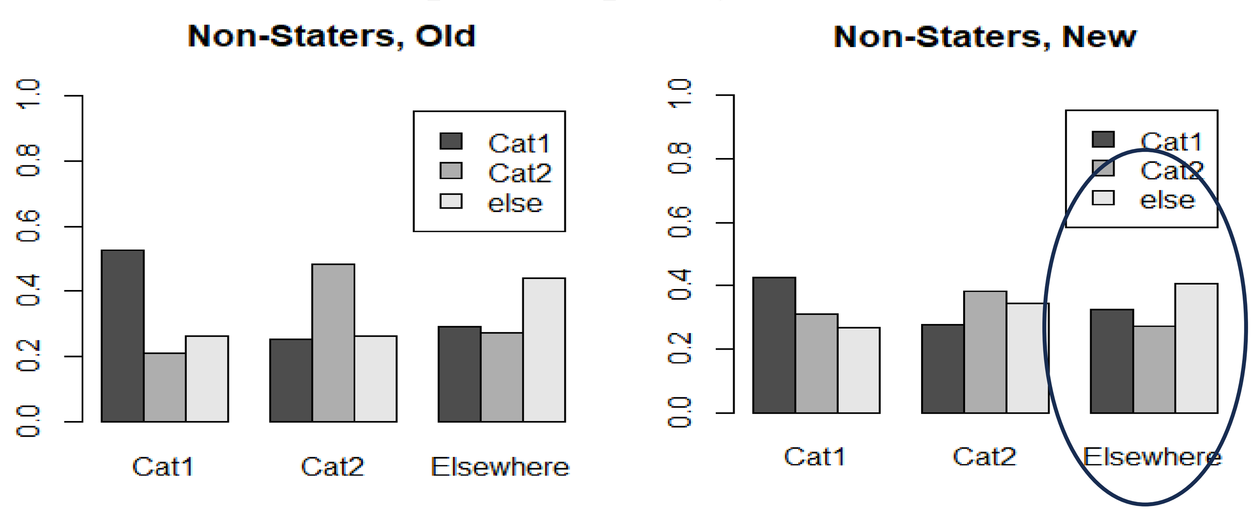


Explicit vs. Implicit learners

Previous work (Moreton & Pertsova, 2016, 2023) shows that ability to verbalize aspects of the experimental pattern after the experiment
- Correlates with measures of implicit/explicit learning (e.g., the speed and bimodality of responses)
- Is potentially correlated with different cognitive biases (coming from System I vs. System II).

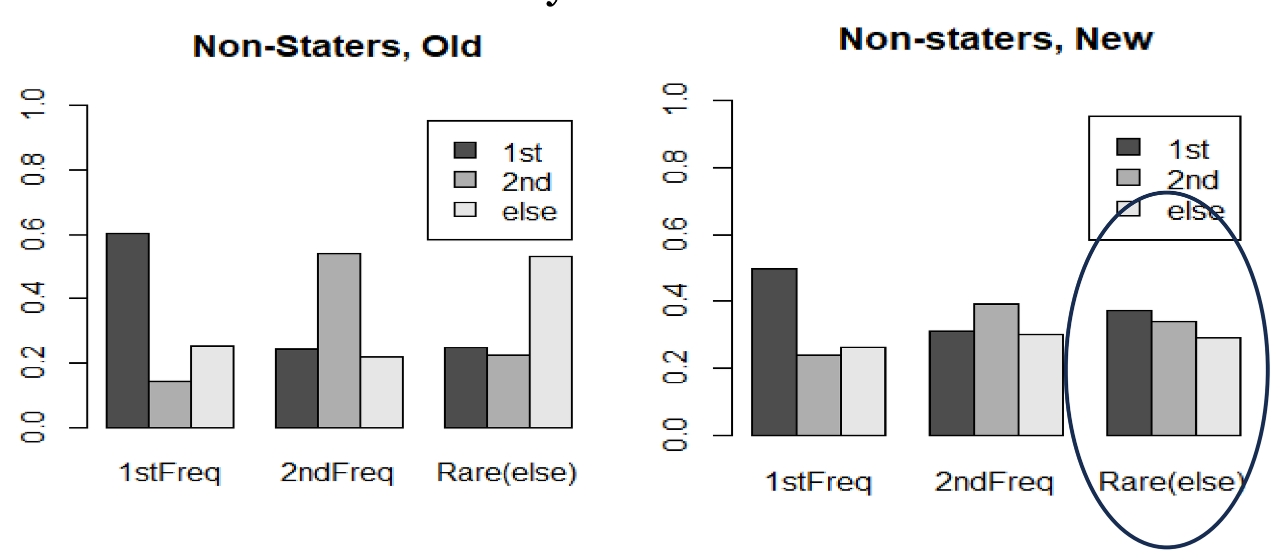
Rule-staters: chose the correct suffix above chance for every category in both equal-frequency and minority-default conditions.
Non-staters: performance varied based on condition and old vs. new stimuli

Equal Frequency Condition



Elsewhere (odds-ratio=1.28, p=0.03)

Minority Default Condition



Most frequent suffix > Elsewhere (odds-ratio=1.28, p=0.03)

Summary of results:
1. In both conditions, the Elsewhere suffix was chosen above chance for Elsewhere category words showing successful acquisition of defaults.
   - However, this was NOT TRUE for non-staters for novel words
2. Frequency had a significant effect in two ways (consistent with Nevat et.al. 2018)
   - improved learners' performance on the narrow categories (stimuli in Group 1 and Group 2 were learned better in minority default condition compared to equal frequency condition).
   - biased non-staters in the minority-default condition towards most-frequent category
3. Speakers ignored syllable count as a relevant feature and focused only on the last segment of the stem (novel trisyllabic words that ended in nasals were treated as Group 1 and those ending in –t -- as Group 2; those who verbalized the rules almost never mentioned the syllable count).

EFFECTS OF FREQUENCY AND DISTRIBUTION ON LEARNING DEFAULTS

K.Pertsova (pertsova@unc.edu), B. Prickett (bprickett@umass.edu), and E.Chen (estherchen@unc.edu)

Encoder-Decoder: Ability to Learn Minority Default

Model: Following past work (Kirov and Cotterell, 2018; McCurdy et al., 2020), we trained an encoder-decoder neural network (Sutskever et al. 2014) to map from stems to their appropriate suffixes. The model:

- Had 2 GRU layers (Cho et al., 2014) in its encoder and decoder
- Used 10 nodes in each layer
- Used hyperbolic tangent activation functions in its output layer

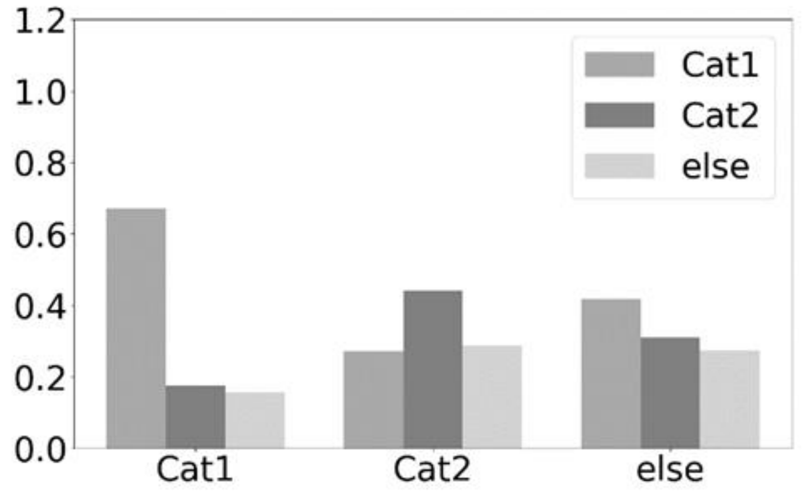
Training: The model learned to map stems to one of the three suffixes (e.g., [bast] → [-wa]) using:

- Standard phonological features representing the input
- Two numerical features representing the output suffixes
- 50 epochs of training and 10 separate runs (with randomly sample initial weights) in each condition

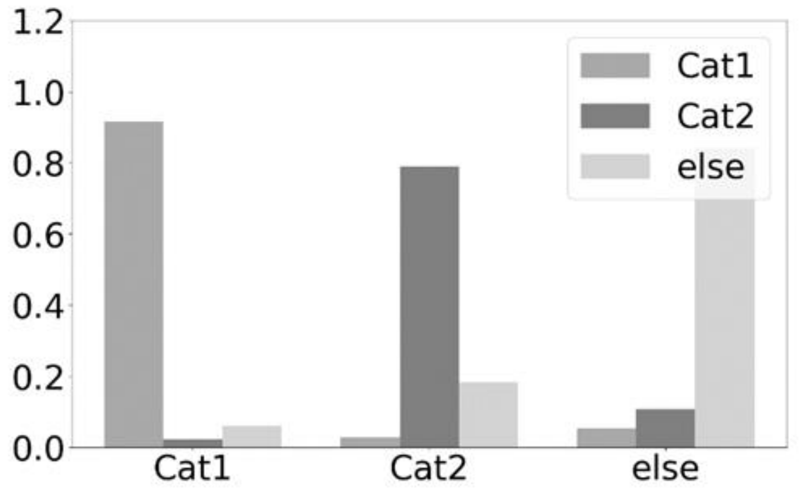
Was the encoder-decoder model able to learn a minority default? When trained on the same kind of data as the human participants in the minority default condition, and tested on novel stems the model:

- Had a frequency effect early in learning (when the model had a similar accuracy to our human participants).
- Successfully captured a minority default later in learning (when the model had converged on the patterns in its training data).

(1) Minority Default: Early in Learning



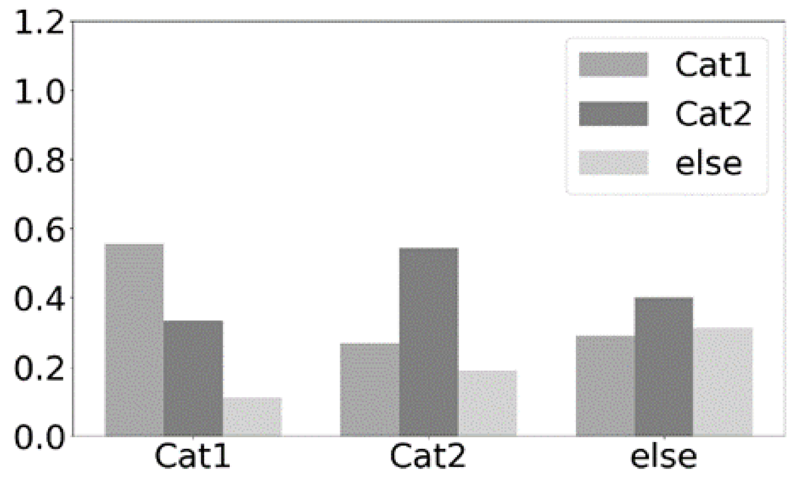
(2) Minority Default: Later in Learning



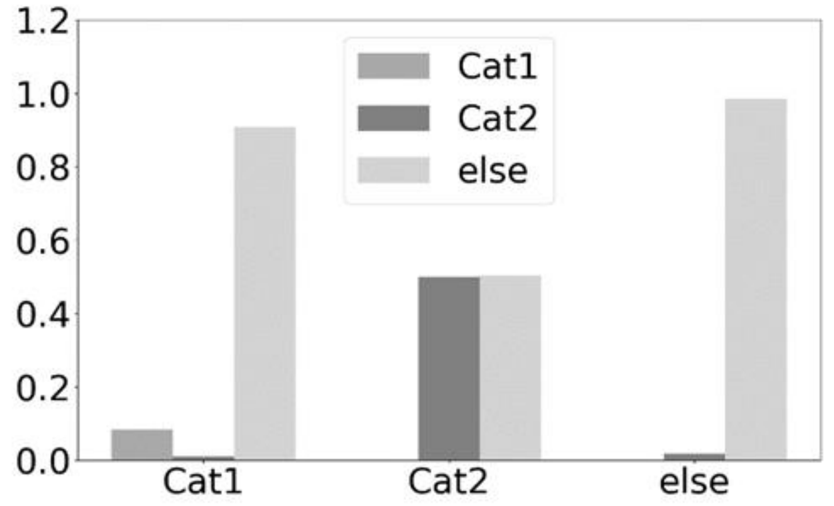
When trained on the "Equal Frequency" condition, and tested on novel stems the model:

- Had no default for "else" stems early in learning.
- And *overgeneralized* the default suffix later in learning.

(3) Equal Frequency: Early in Learning



(4) Equal Frequency: Later in Learning



Encoder-Decoder: Sensitivity to Syllable Number

Syllable count: Experiment participants were tested on words that:

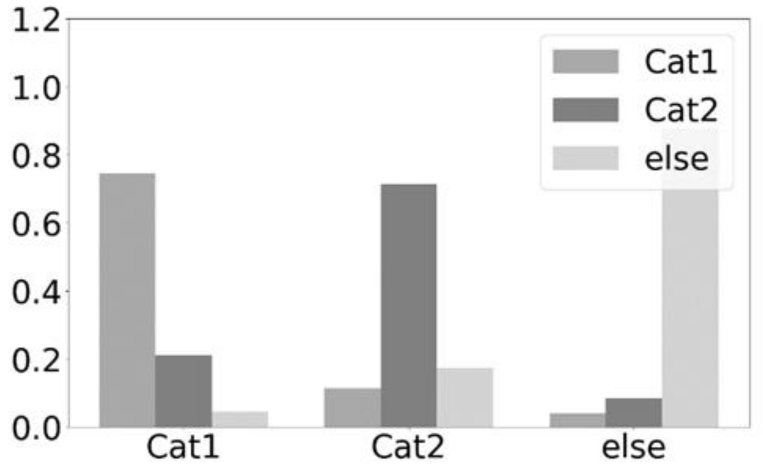
- Were novel and matched the number of syllables in training for each category
- And novel words with three syllables

The number of syllables in test items, however, didn't significantly affect human generalization.

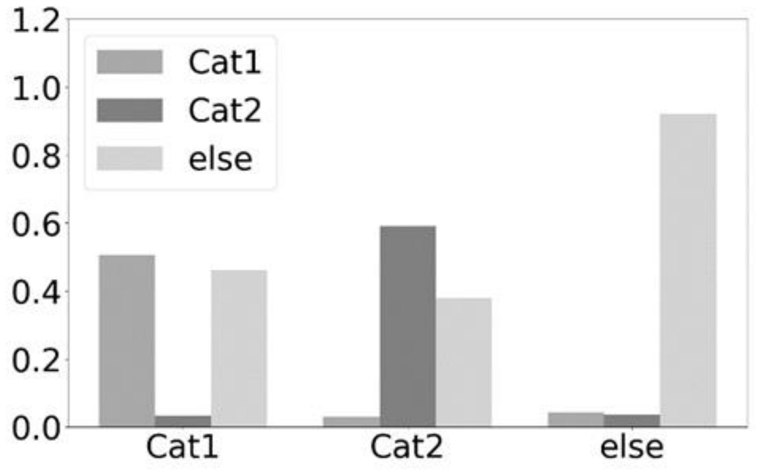
Was the encoder decoder model sensitive to syllable count?
The results in Figure (1-4) only show the model's generalization to novel stems with 1 or 2 syllables, as in training). Figures 5 & 6 show the results on trisyllabic stems

- Early in learning, there was no notable differences between these results and how the model generalized to words with novel syllable counts.
- However, later in learning, the model behaves very differently on words with novel syllable counts (results below are for 3-syllable stems):

(5) Equal Freq: Novel Syllable Count



(6) Min. Default: Novel Syllable Count



Feedforward Network: No Minority Default

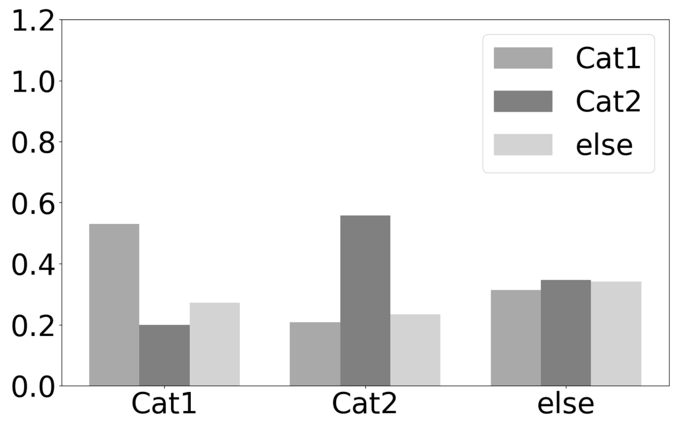
Model: Work that originally tested neural networks' ability to learn minority defaults used a feed forward architecture (Hare et al. 1995).

- We tested this architecture as well, with a model that had two hidden layers.
- There were 50 nodes in each layer, each with a bias connection

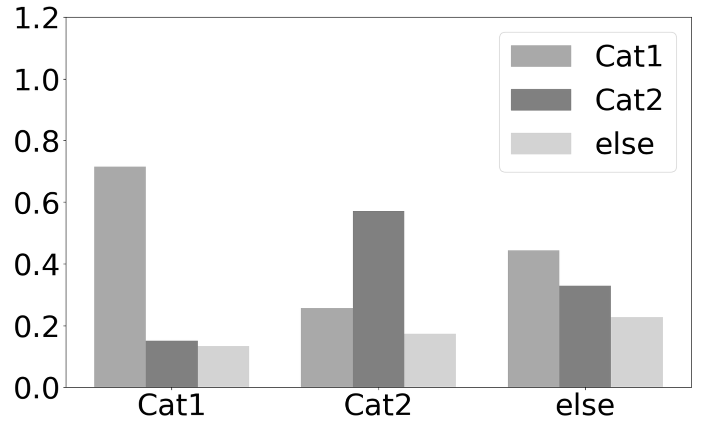
Results: When trained on the same kind of data as the human participants:

- We found that this model failed to learn a minority default and was solely affected by the frequency of each category.

(7) Equal Frequency: Feed Foward



(8) Minority Default: Feed Foward

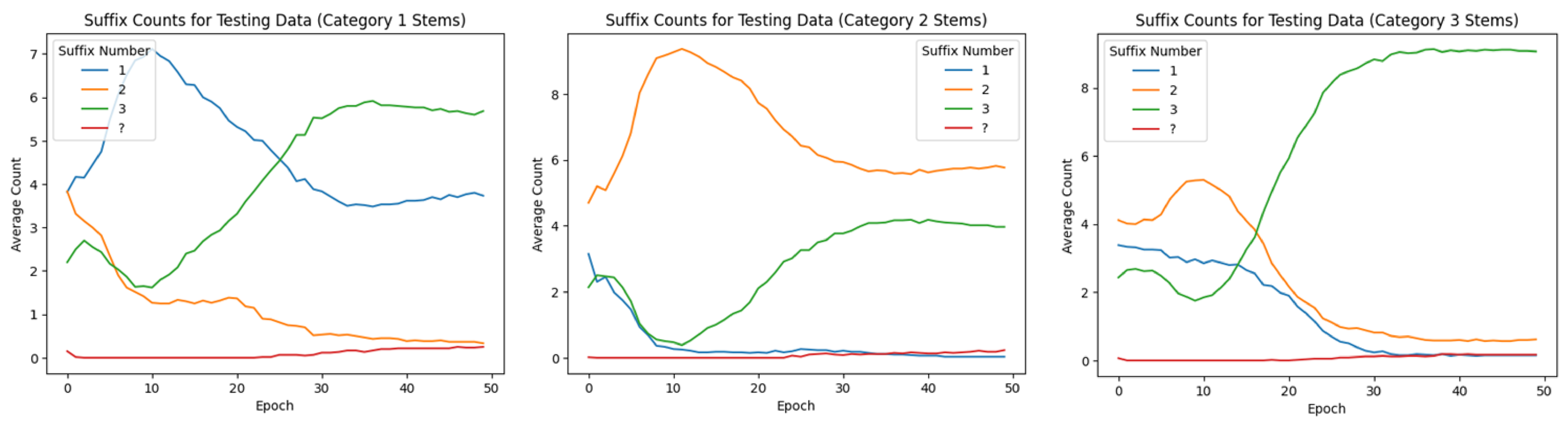


Encoder-Decoder: Generalization Over Time

The outcome for the network is dependent on when in learning we ask it to generalize: So how does that change over time progress?

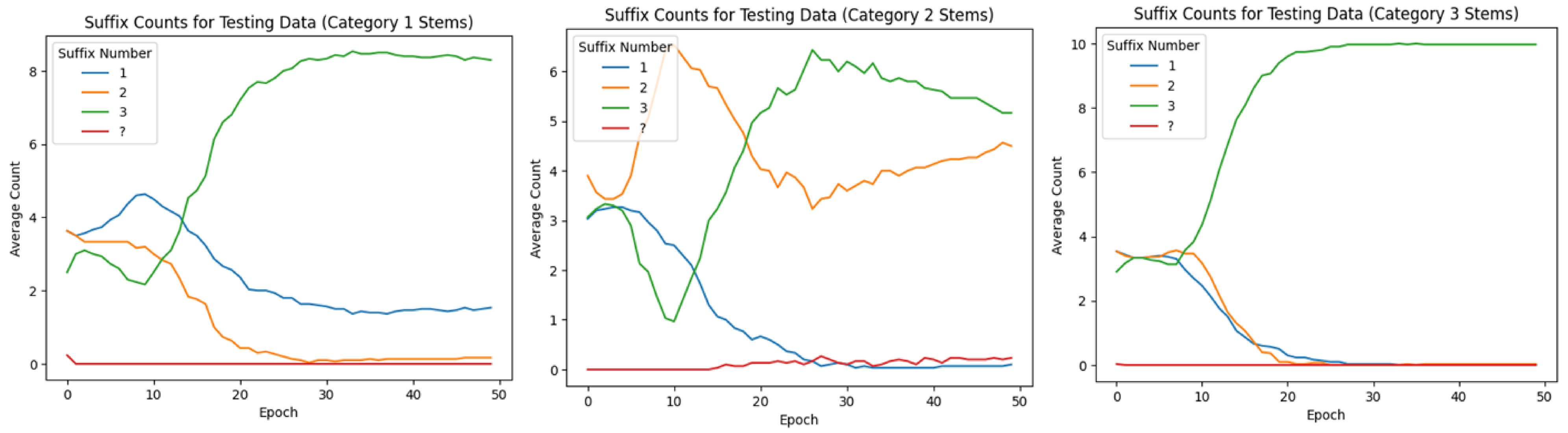
- In the minority default condition, the model's output for stems in category 1 changes around epoch 25.
- That same condition sees the model change its output for stems in category 3 a bit earlier—around epoch 20.

(9) Minority Default: Generalization to Novel Stems over Time



- For the equal frequency condition, stems belonging to categories 1 and 2 change around epoch 20.
- But stems belonging to category 3 change earlier—around epoch 10.

(10) Equal Frequency: Generalization to Novel Stems Over Time



Conclusions

Question 1: Are minority-defaults learnable?

- YES (even given relatively short exposure), for explicit (or fast) learners. Implicit learners possibly need more time.
- YES, but only under some conditions, Strangely, an ED learner struggles on the regular (non-minority) default, by overgeneralizing it.

Question 2: Does frequency trump distribution?

- YES for implicit leaners
- YES for ED model at the early stages of learning

Post-hoc finding

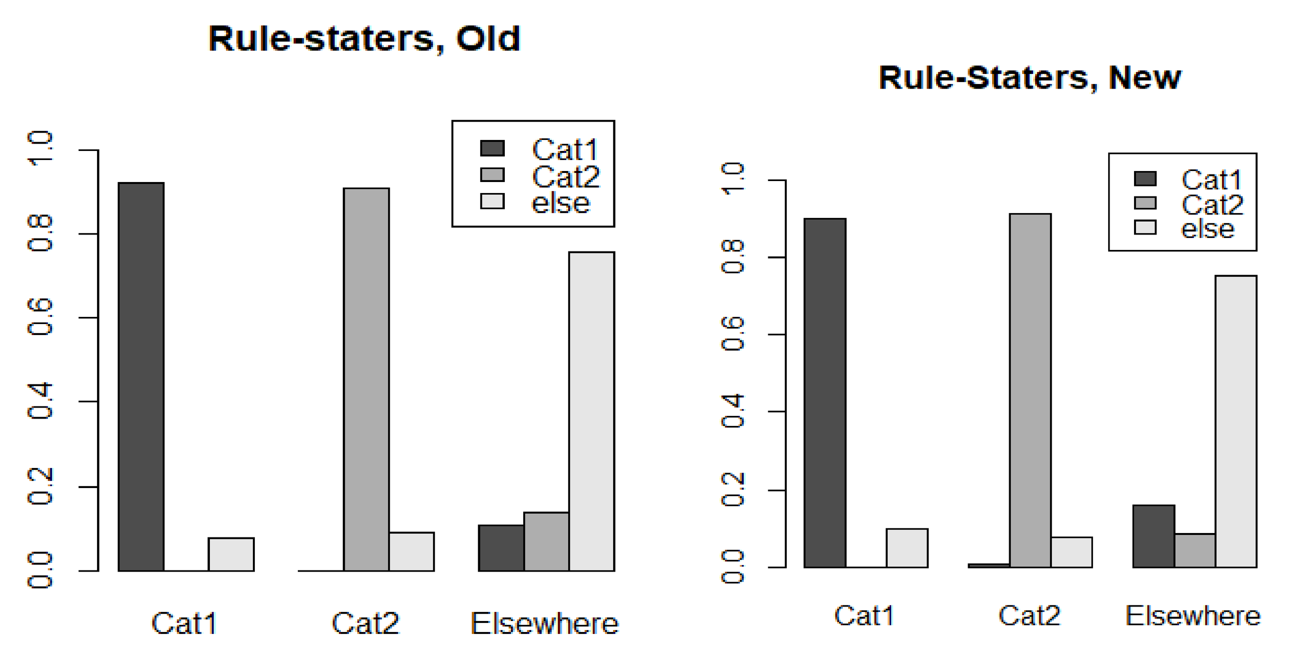
- People generalized based on the immediately adjacent to the suffix final consonant of the stem only (ignoring syllable number or clusters)

EFFECTS OF FREQUENCY AND DISTRIBUTION ON LEARNING DEFAULTS

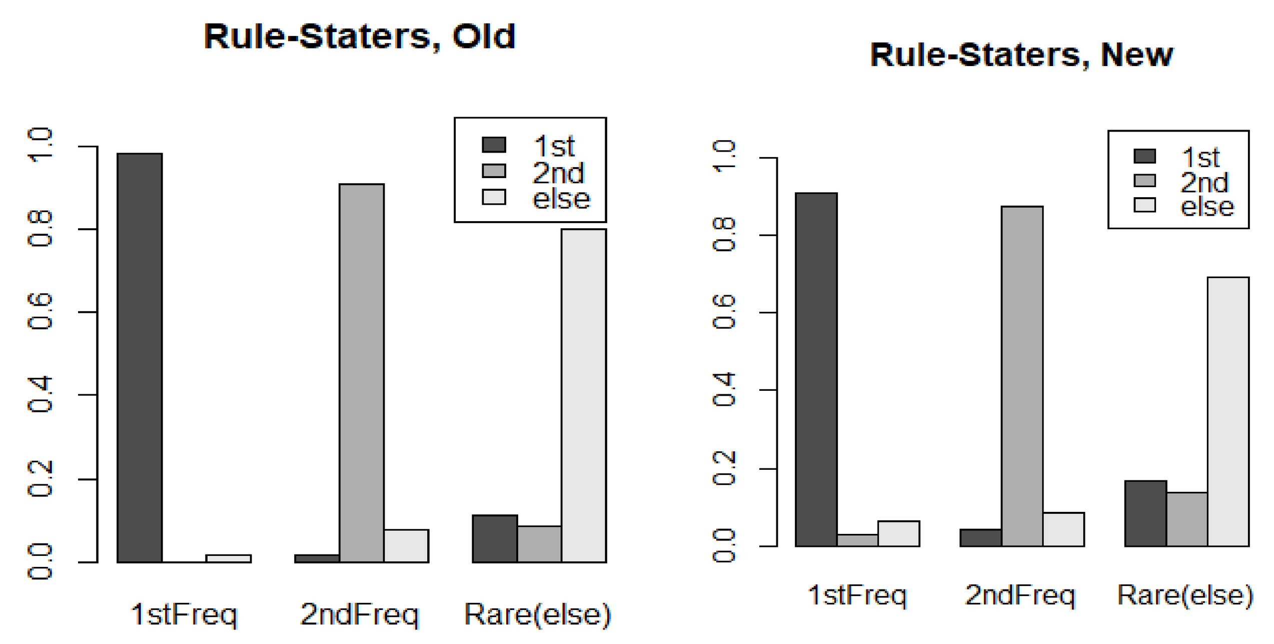
K.Pertsova (pertsova@unc.edu), B. Prickett (bprickett@umass.edu), and E.Chen (estherchen@unc.edu)

Supplementary graphs

Equal Frequency Condition



Minority Default Condition



Proportion of suffix choices per category for rule-staters

References:

Aronoff, M. (1976). Word formation in generative grammar. *Linguistic Inquiry Monographs Cambridge, Mass*, (1), 1-134.

Boudelaa, S., & Gaskell, M. G. (2000). In search of the minority default: the case of Arabic plurals. In *Proceedings of CogSci 2000* (pp. 48-53).

Forrester, N., & Plunkett, K. (2019). Learning the Arabic Plural: The Case for Minority Default Mappings in Connectionist Networks. In *Proceedings of the sixteenth annual conference of the Cognitive Science Society* (pp. 319-323). Routledge.

Cho, K., Van Merriënboer, B., Bahdanau, D., & Bengio, Y. (2014). On the properties of neural machine translation: Encoder-decoder approaches. *arXiv preprint arXiv:1409.1259*.

Hahn, U., & Nakisa, R. C. (2000). German inflection: Single route or dual route?. *Cognitive Psychology*, *41*(4), 313-360.

Hare M., J. Elman & Kim Daugherty (1995). Default generalization in connectionist networks. *Language and Cognitive Processes*, 10 (1995), pp. 601-630

Kirov, C., & Cotterell, R. (2018). Recurrent neural networks in linguistic theory: Revisiting Pinker and Prince (1988) and the past tense debate. *Transactions of the Association for Computational Linguistics*, *6*, 651-665.

Kiparsky, P. (I973) "'Elsewhere' in Phonology," in S. R. Anderson and P. Kiparsky, eds., *Festschrift for Morris Halle*, Holt, Rinehart and Winston, Inc., New York.

McCarthy, J. J., & Prince, A. S. (1990). Foot and word in prosodic morphology: The Arabic broken plural. *Natural Language & Linguistic Theory*, *8*(2), 209-283.

Marcus, G. F., U. Brinkmann, H. Clahsen, R. Wiese, & S. Pinker. (1995). German Inflection: The Exception That Proves the Rule. *Cognitive Psychology, 29*(3), 189-256.

McCurdy (2019). Neural Networks Don't Learn Default Rules for German Plurals, but that's Okey, neither do Germans. MA Thesis. U. Of Edingburgh.

McCurdy, K., Goldwater, S., & Lopez, A. (2020). Inflecting when there's no majority: limitations of encoder-decoder neural networks as cognitive models for German plurals. *arXiv preprint arXiv:2005.08826*.

Moreton, E., & Pertsova, K. (2016). Implicit and explicit processes in phonotactic learning. In *Proceedings of the 40th Boston University Conference on Language Development, Somerville, Mass., pp. TBA. Cascadilla*.

Moreton, E., & Pertsova, K. (forthcoming). Implicit and explicit processes in phonological concept learninog. *Phonology*.

Nevat, M., Ullman, M. T., Eviatar, Z., & Bitan, T. (2018). The role of distributional factors in learning and generalising affixal plural inflection: An artificial language study. *Language, Cognition and Neuroscience*, *33*(9), 1184-1204.

Pinker, S., & Ullman, M. T. (2002). The past and future of the past tense. *Trends in Cognitive Sciences*, 6(11), 456–463.

Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. *Advances in neural information processing systems*, *27*.

Zaretsky, E., Müller, H. H., & Lange, B. P. (2016). No default plural marker in Modern High German. I*talian Journal of Linguistics*, *28(*1), 1-28.