Effects of frequency & distribution on learning defaults

Some linguistic patterns (e.g., English past tense regular inflection) are often stated using *default* rules, i.e., general rules that apply as a last resort. How such patterns are learned has been a subject of a long and intense debate with particular focus on *minority defaults* and the ability of neural nets to learn them (Hare et al., 1995; Pinker and Ullman, 2002). Recent work by McCurdy et al. (2020), building on a model of Kirov and Cotterell (2018), tests an Encoder-Decoder (ED) model on the problem of learning minority defaults using data from German plural formation. They found that while their model achieved 88.8% accuracy on the held-out data, it did not match human performance in a few critical ways. In particular, unlike humans, it failed to generalize the two apparent default suffixes -en and -s (where -s is the least frequent suffix) to unusual novel words. The questions we seek to address are (i) do humans in fact rely on default rules/patterns, (ii) what factors influence the development of defaults, and (iii) can ED models successfully learn them?

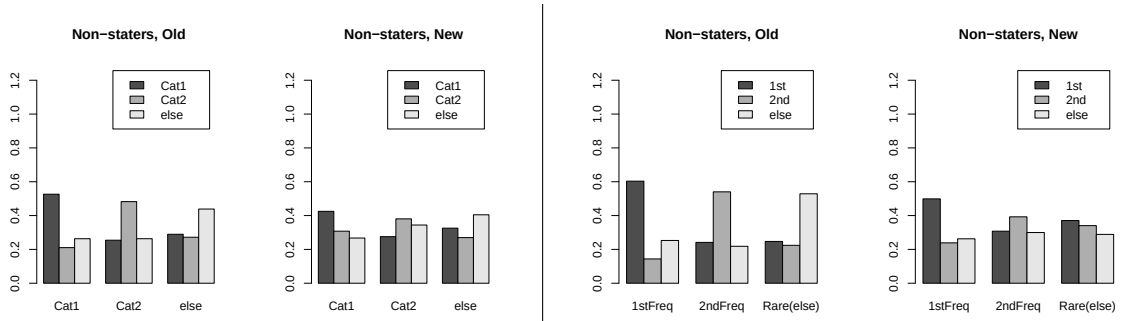
Experiment: Instead of looking at data from real languages like German, in which the default status of the minority suffix is controversial (Zaretsky and Lange, 2015), we test humans and an ED model on an artificial language. In this language, there are three plural allomorphs that depend on the phonological properties of the stem, such that one of the suffixes has a heterogeneous elsewhere distribution (the default), unlike the distribution of the other two, narrowly defined suffixes (see Table 1). We test whether the distribution of suffixes is learned correctly when the frequency of all suffixes is equal (equal frequency condition) and when the default is the least frequent category (minority-default condition).

Procedure: Participants learned how to pluralize novel words in an artificial language in an online experiment. In the first phase, participants were presented with the audio-picture pairs for a singular form, followed by the plural form. In the second phase, they were presented with the same pairs, but had to complete a forced-choice-task with correct/incorrect feedback. The test phase also involved a forced-choice task and included both trained and new instances. A debrief questionnaire then asked participants to describe what they learned.

Table 1: Distribution of three suffixes

	Cat1 (narrow)	Cat2 (narrow)	Cat3 (elsewhere)
Features	2syll + -N	1syll + -Ct	other
Examples	*ranom, cotin, pashem*	*boft, lunt, frest*	*trofa, nasp, sopis*

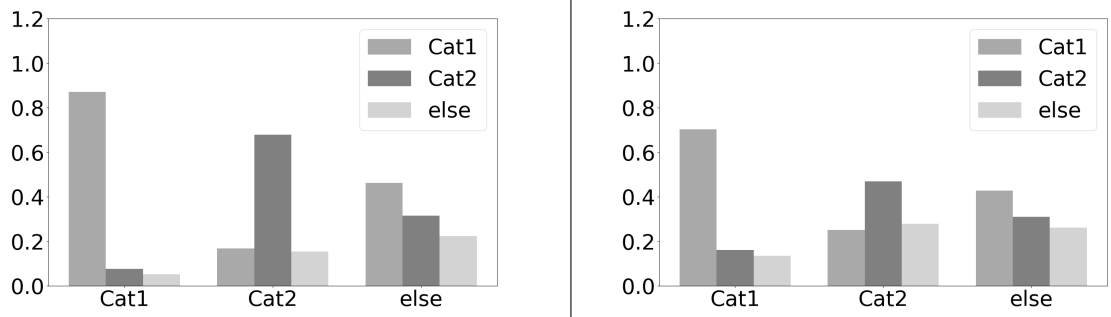
Results: Participants in both conditions were able to correctly generalize suffixes to novel words, including the default suffix. However, they ignored the correlated feature of syllable number and made their decisions based on the final segment of the stem (-t vs. -N vs. elsewhere), suggesting a strategy that focuses on the simplest hypothesis (a single most salient feature). Once we separated participants into those who were able to state at least a partially correct rule (rule-staters) and those who did not (non-staters), an interesting pattern emerged. Rule-staters performed similarly in the two conditions and on old vs. new items, and generally much better across the board compared to non-staters. Non-staters in the equal frequency condition (Fig.1 a) numerically preferred the correct suffixes for both old and new items including the default environment. In a multinomial logistic regression, the preference for the default suffix, given an elsewhere novel stem, was not quite significant



(a) Equal frequency

(b) Minority Default

Figure 1: % of choices of the 3 suffixes in 3 conditions for the humans.



(a) Stems from Training

(b) Novel Stems

Figure 2: % of choices of the 3 suffixes in 3 conditions for the ED model, early in learning.

relative to Cat.1 (odds = 1.13, p=0.1), while it was significant relative to Cat.2 suffix (odds = 1.15, $p < 0.001$). However, in the minority-default condition (Fig. 2b), non-staters preferred one of the more frequent suffixes in the novel default context, with the preference for the most frequent suffix being significant (odds = 1.28, p=0.03). This suggests that in early stages of learning or when learning is implicit, learners tend to treat the majority pattern as the default. This is consistent with findings of Nevat et al. (2018). When frequency doesn't play a role or learning reaches conscious awareness, it is the elsewhere suffix that becomes the default.

Model: We trained an ED model with 2 GRU layers (Cho et al., 2014) in its encoder and decoder for ten separate repetitions in each condition using the learning algorithm Adam (Kingma and Ba, 2014). The model learned to map stems (which were represented as a sequence of phonemes) to one of the three suffixes. When trained on the same kind of data as the human participants, the model shows a frequency bias at early stages of learning (Fig.2), even for words present in its training data (unlike humans). Later in learning, the model learns the elsewhere category but tends to overgeneralize it to all new words, especially in the equal-frequency condition, which also fails to mirror the behavior of our human participants. These results suggest that frequency trumps distribution early in learning, but in the end humans *can* learn minority defaults, while the network overgeneralizes them.