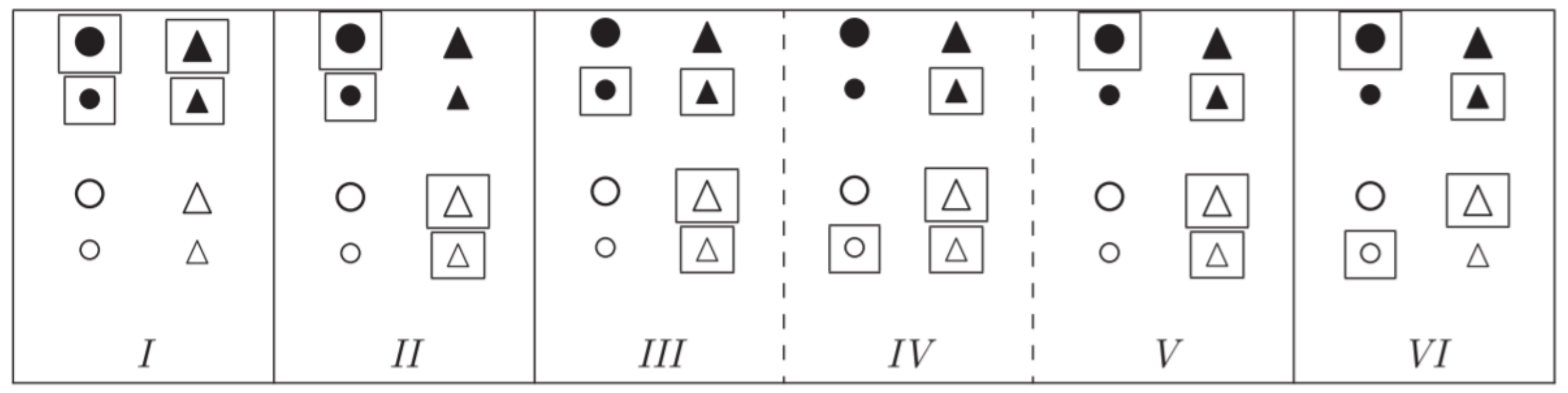
Capturing Phonotactic Learning Biases with a Simple RNN

Max Nelson*
Joe Pater
Brandon Prickett
*manelson@umass.edu
CMCL 2021

Introduction

- Shepard et al. (1961) define six pattern types that represent the possible ways of dividing a space defined by three binary features in half
- See the visual space below, defined by [±circle], [±black], and [±small]

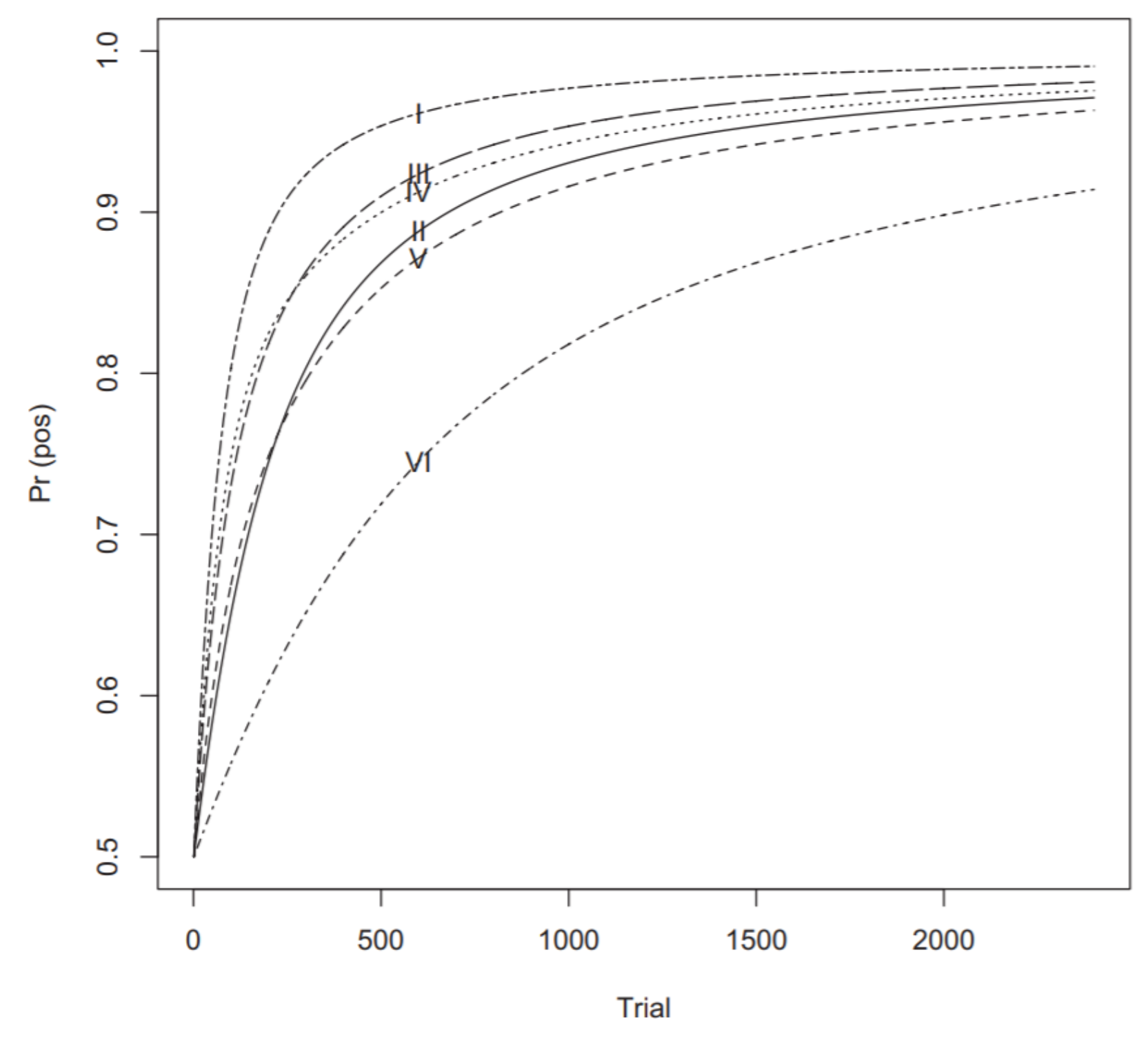


- Moreton et al. (2017) - artificial language learning implementing Shepard type patterns phonologically in CVCV nonce words
  - Consonant features: [±voice], [±coronal] (for first and second consonant)
  - Vowel features: [±high], [±back] (for first and second vowel)
  - e.g. “Only [+voice] consonants can begin a word” (Type I)
  - Report the following ranking of the Shepard type patterns in order of difficulty of human learning:

Humans: *I* > *IV* > *III* > *V* > *II* > *V*

- Moreton et al. (2017) used a MaxEnt grammar with an unbiased constraint set - every possible conjunction of features
  - e.g. $*[+\text{voice}]_{C1}[-\text{voice}]_{C2}$ - Violated if first consonant is [+voice] and the second is [-voice]

MaxEnt Model: *I* > *III*, *IV* > *II*, *V* > *VI*



- With an early *IV* > *III* and *V* > *II* biases that later reverse
- MaxEnt models like Moreton et al. require explicit negative evidence
- Real language learners, and those participating in Moreton et al.'s study, do not have labelled examples of ungrammatical strings

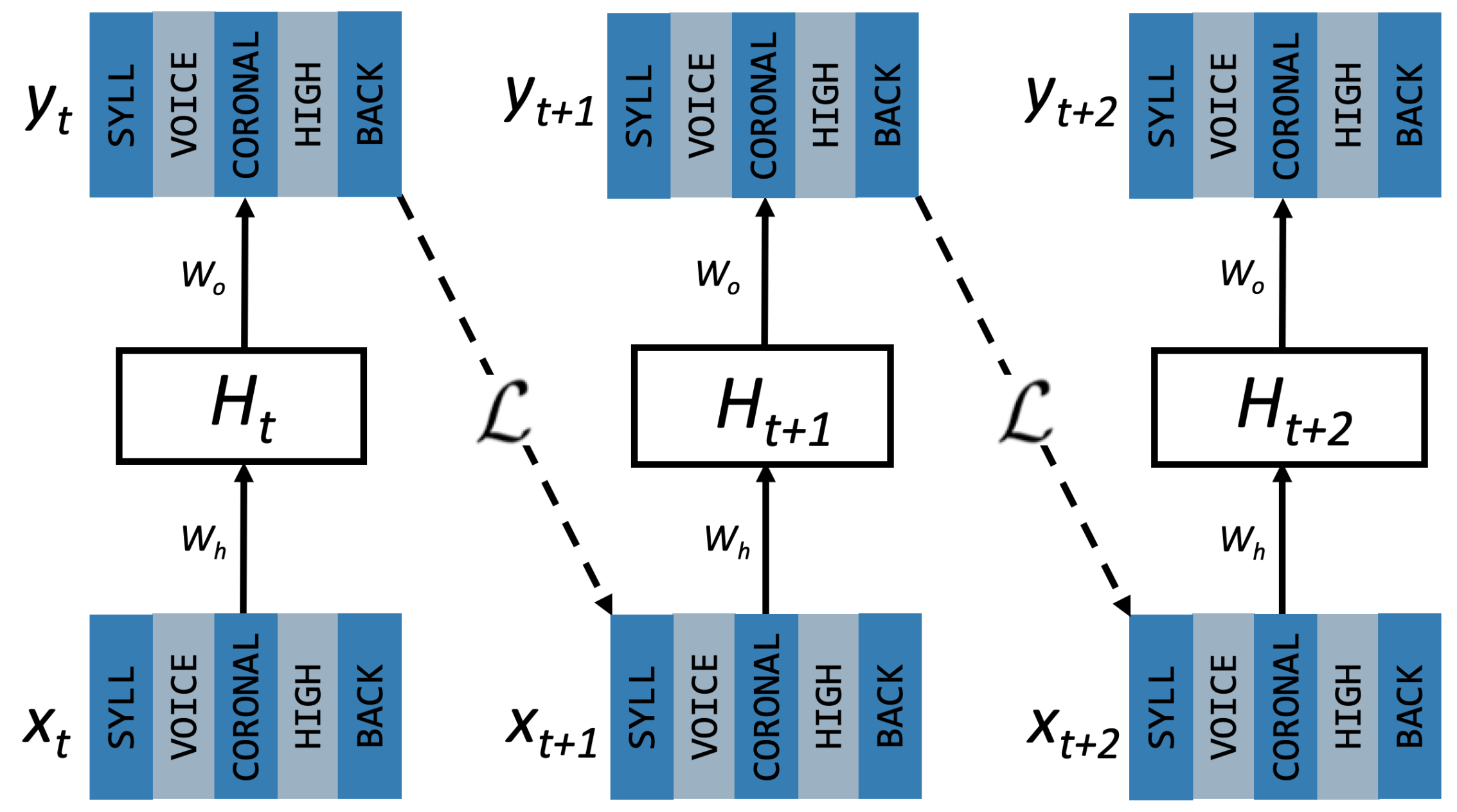
We test if RNNs trained on a language modeling objective with no negative evidence exhibit the same biases observed by Moreton et al. (2017)

RNNs for phonotactic learning

- Hare (1990) and Gasser and Lee (1992) first applied RNNs to phonology.
- Since then, RNNs (Doucette, 2017) and RNN language models more specifically (Rodd, 1997; Mirea and Bicknell, 2019; Mayer and Nelson, 2020) have been shown to be viable approaches to modeling phonotactic learning
- Language modeling is a useful way of capturing phonotactic learning because:
  - It doesn't need any negative evidence (i.e., it's *unsupervised*)...
  - ...And because RNNs have a limited amount of built-in linguistic structure.

Methods

- Our networks are simple RNN language models - inputs and outputs are binary feature vectors
  - Sigmoid output layer, binary cross-entropy loss, weights optimized using Adam (Kingma and Ba, 2014)



- Our networks are simple RNN language models - inputs and outputs are binary feature vectors

	syllabic	*voice*	*coronal*	*high*	*back*
t	−	−	+	−	−
k	−	−	−	−	−
d	−	+	+	−	−
g	−	+	−	−	−
i	+	−	−	+	−
u	+	−	−	+	+
æ	+	−	−	−	−
a	+	−	−	−	+

- We trained this model on words that were randomly constructed in the exact same manner as the stimuli in Moreton et al.'s (2017) training phase.
  1. First, a Shepard type was chosen (e.g., Type I).
  2. Then the relevant feature(s) for the pattern were chosen as well as the value(s) for the feature(s) that would define whether a word was grammatical or not (e.g., “first consonant must be [+voice]”).
  3. The full space of possible stimuli were then split based on whether they were grammatical or not.
  4. 32 items were sampled without replacement from the set of grammatical items and these were used as the model's training data.
- Testing data for the model was also created using the same method that Moreton et al. (2017) used in their experiment:
  1. Out of the grammatical and ungrammatical items created when constructing the training data, 32 novel grammatical items (i.e., words that were not present in training) and 32 ungrammatical items were randomly sampled without replacement.
  2. Grammatical and ungrammatical words were randomly paired to simulate the forced choice given to Moreton et al.'s (2017) participants.

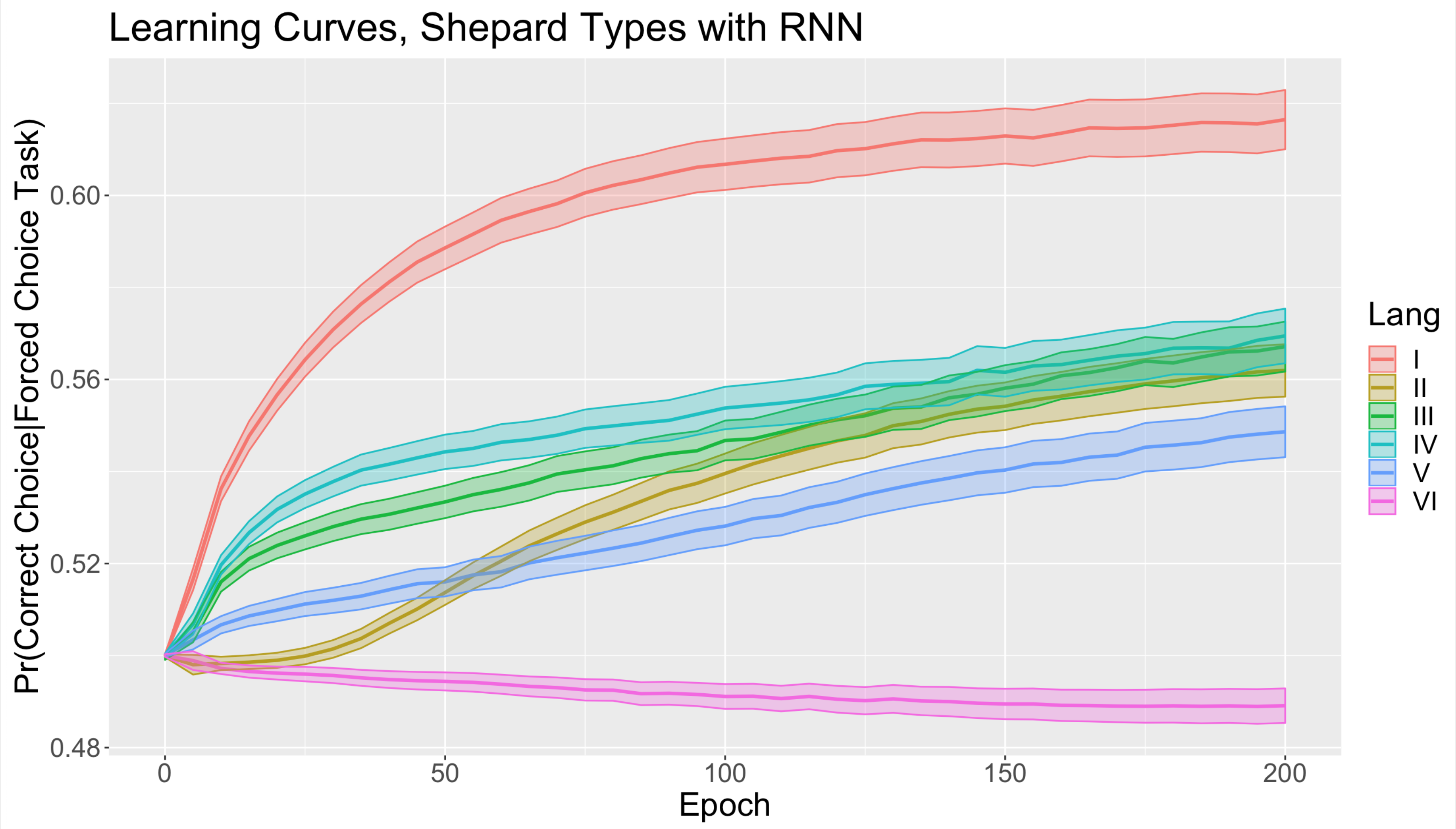
Results

- At each epoch of learning, we recorded the model's error on the test data and estimated probabilities for each word using this error:

$$p(w_1 \mid w_1, w_2) = 1 - \frac{\mathcal{L}(w_1)}{\mathcal{L}(w_1) + \mathcal{L}(w_2)}$$

- The figure below shows these probability estimates over the course of 200 epochs of training:

RNN Model: *I* > *IV* > *III* > *II*, *V* > *VI*



Future Work

- Future work should investigate whether the phonological features given to our model in its input and output are necessary or if learned features could also capture these biases (as in, e.g., Mayer and Nelson, 2020).
- It could also test this model on other biases that have been observed in artificial language learning, such as *Intradimensional Bias* (Moreton, 2012).
- Additionally, exploring how well this model scales up to real language data could be useful for cases observed in natural language that have been used to argue against Connectionist accounts of phonology (see, e.g., Berent, 2013).

Discussion

- We found that the RNN language model could capture the biases observed in the Moreton et al. (2017) experiment.
- It did this with **no negative evidence** and **no *a priori* constraints**.
- This suggests that prespecified constraint sets are not necessary for a model to capture these results, casting doubt on whether innate constraints play a role in human phonotactic learning.