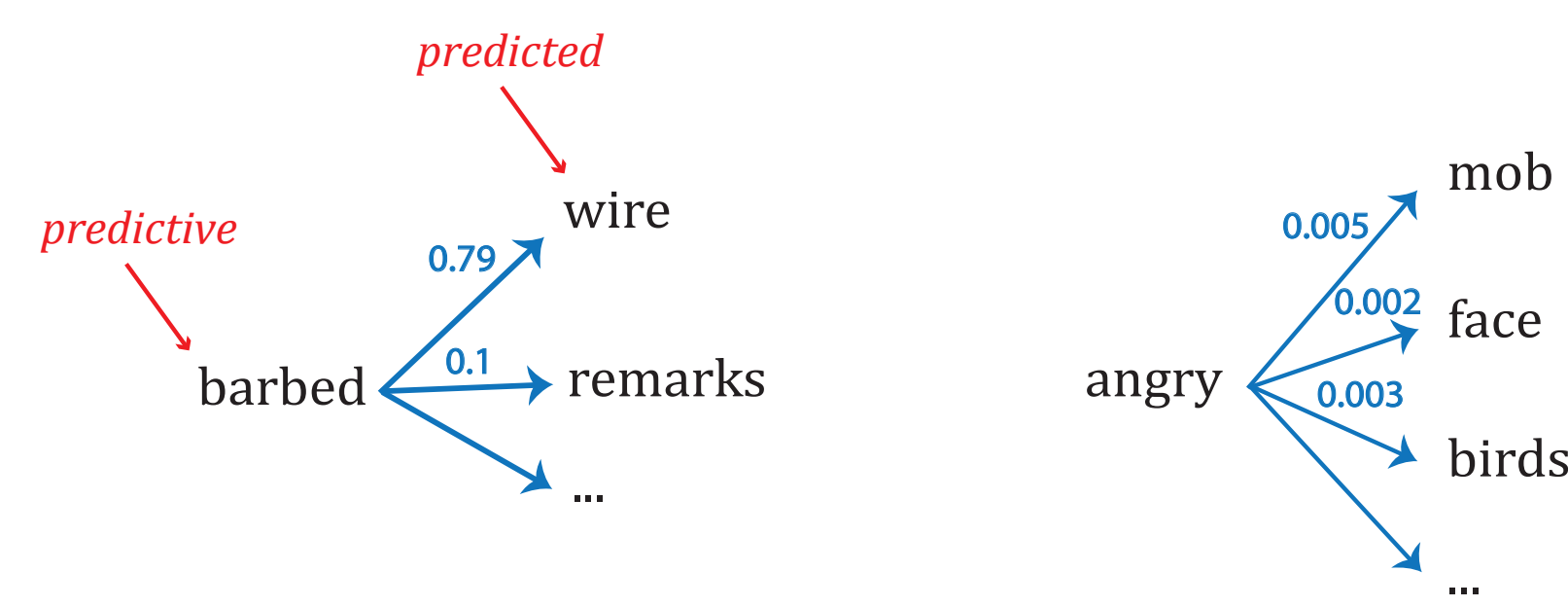


Prediction and uncertainty in an artificial language

Tal Linzen¹, Noam Siegelman² and Louisa Bogaerts²

¹Johns Hopkins University ²Hebrew University of Jerusalem

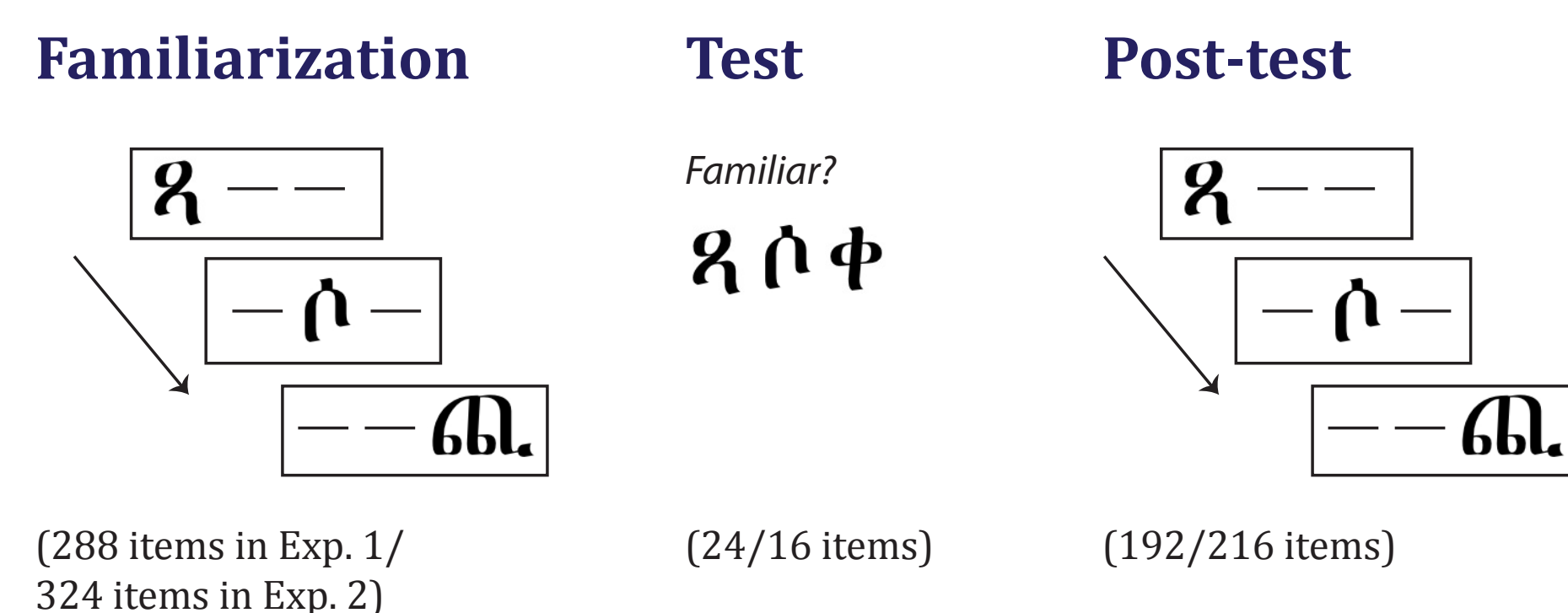
Motivation: prediction in language



- Factors that may affect predictive processing (see Linzen, 2015):
 - The **predictability** of the predicted word (surprisal: $-\log_2 p_i$)
 - The **entropy** (over predictions) at the predictive word: $-\sum_i p_i \log_2 p_i$
 - Entropy reduction** between the predictive and predicted word (disambiguation cost)
- Difficult to control probabilities (especially entropy) in natural language
- Self-paced reading (Just et al., 1982) shows predictability effects, even in an artificial language (Karuza et al., 2014)
- We construct languages that have the desired probability distributions

Artificial language study: methods

- Sequences of three Ge'ez glyphs (first glyph is fixed)
- Design:**

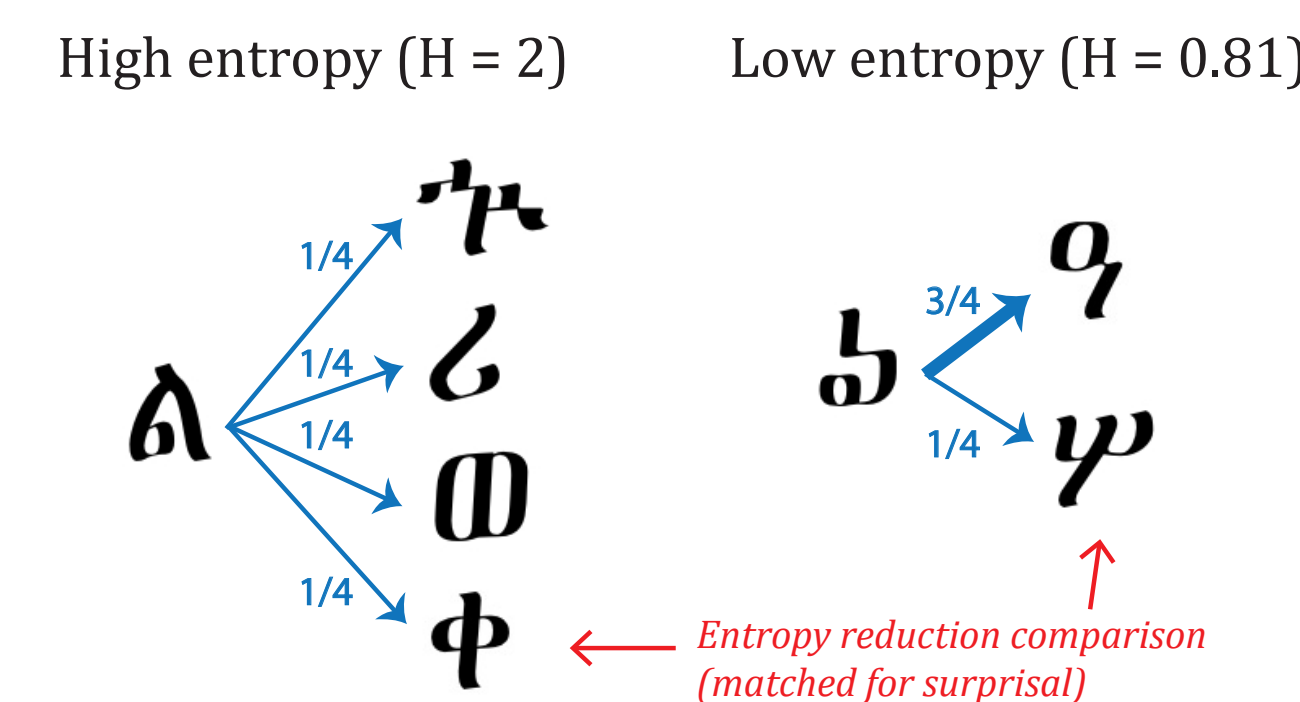


- Only participants whose accuracy was "significantly" higher than chance included in the analysis (23 out of 44 in Exp. 1; 33 out of 46 in Exp. 2)
- Glyphs counterbalanced across participants

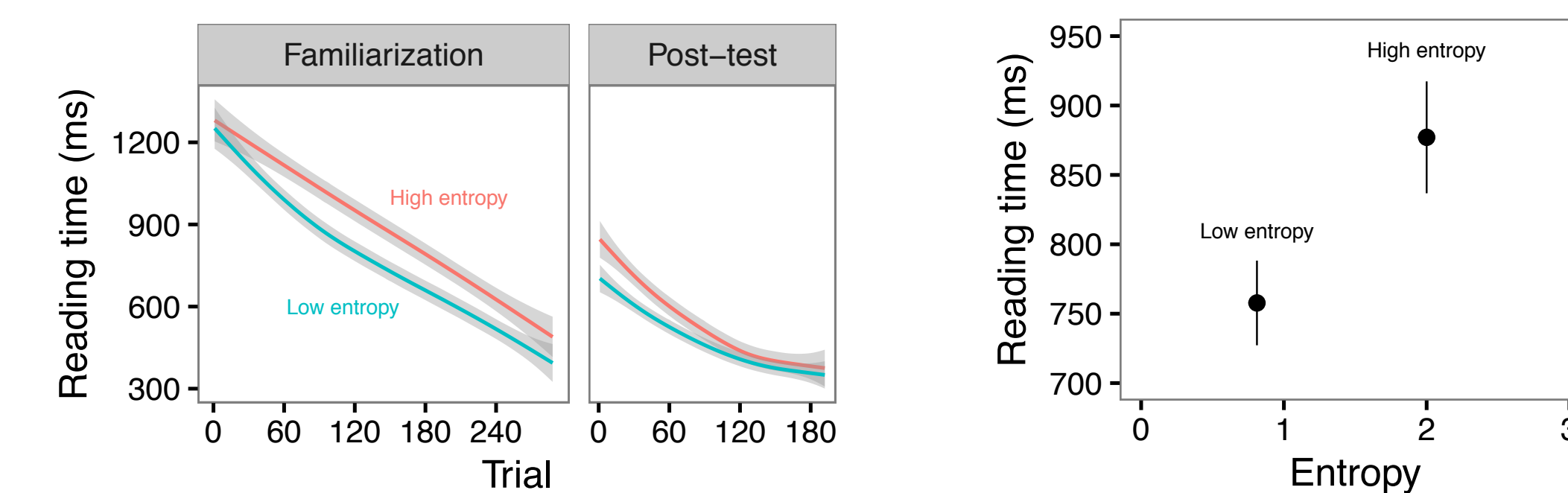
Experiment 1

Goals

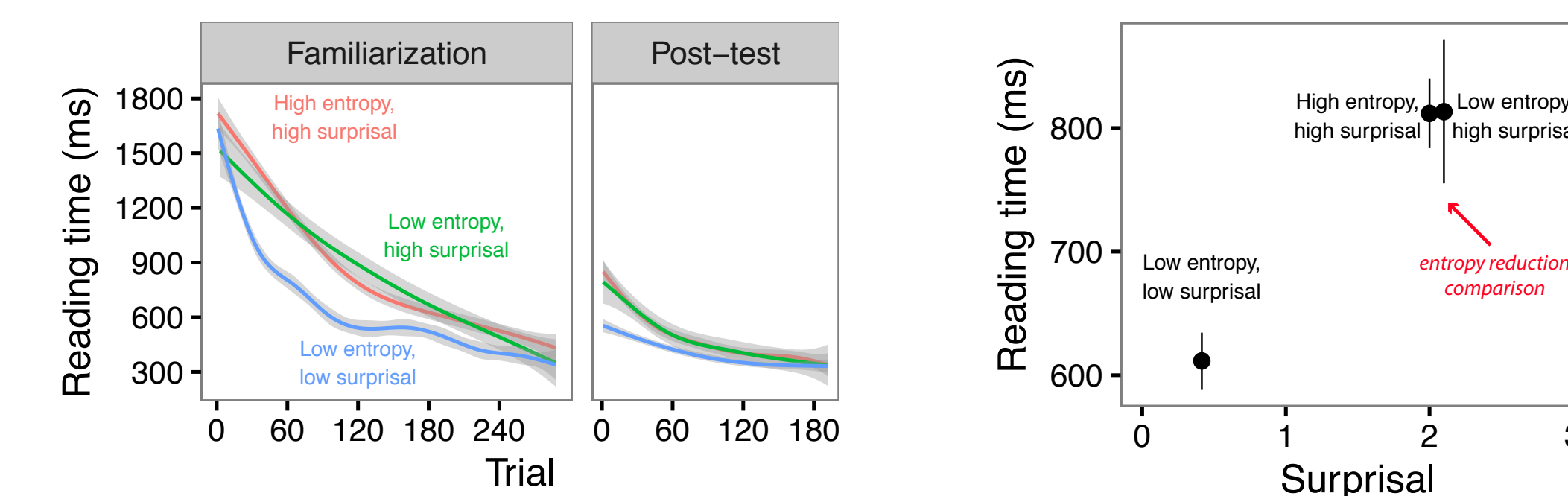
- Surprisal at predicted
- Entropy at predictive
- Entropy reduction at predicted



Predictive glyph



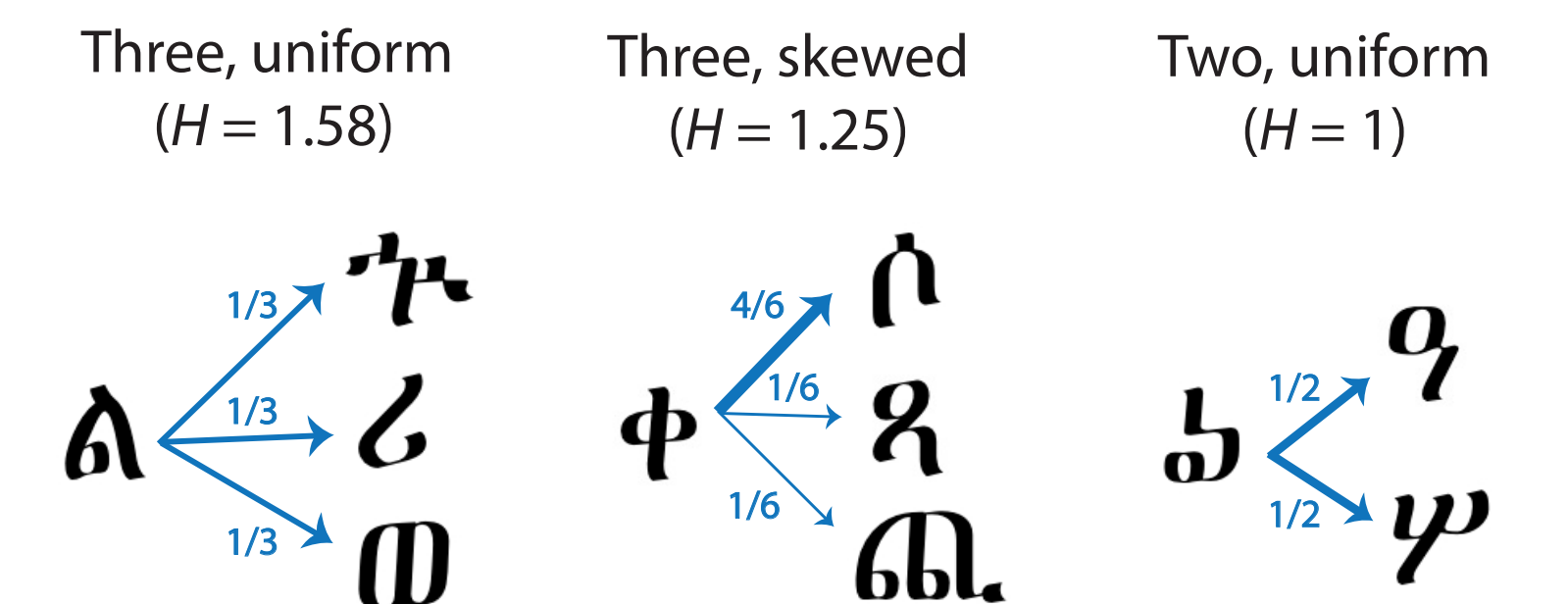
Predicted glyph



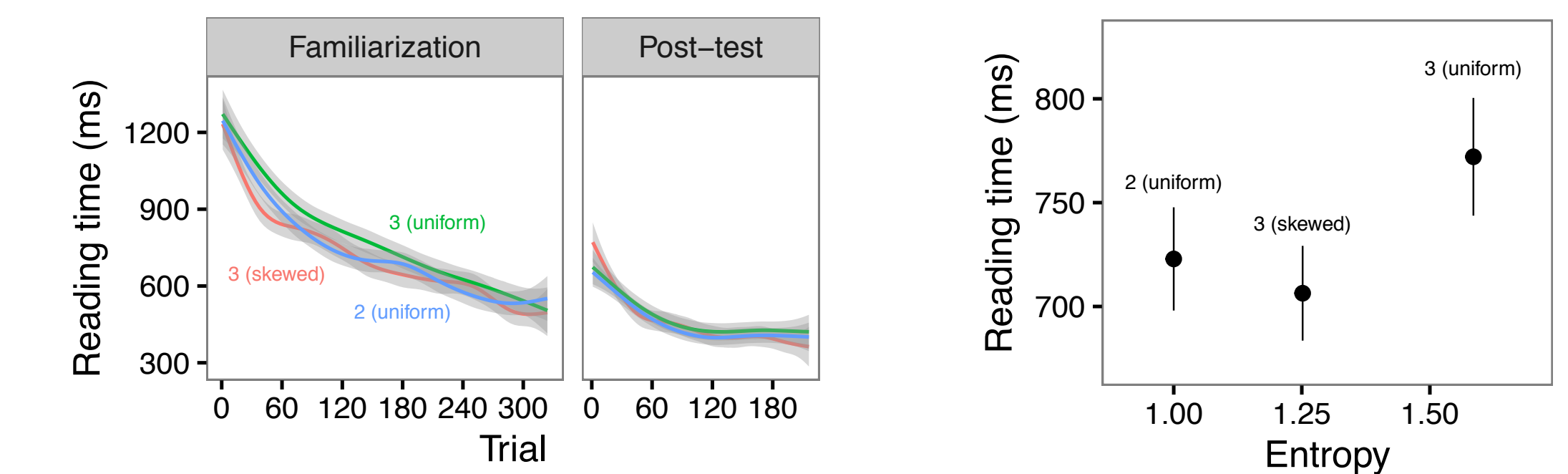
Experiment 2

Goals

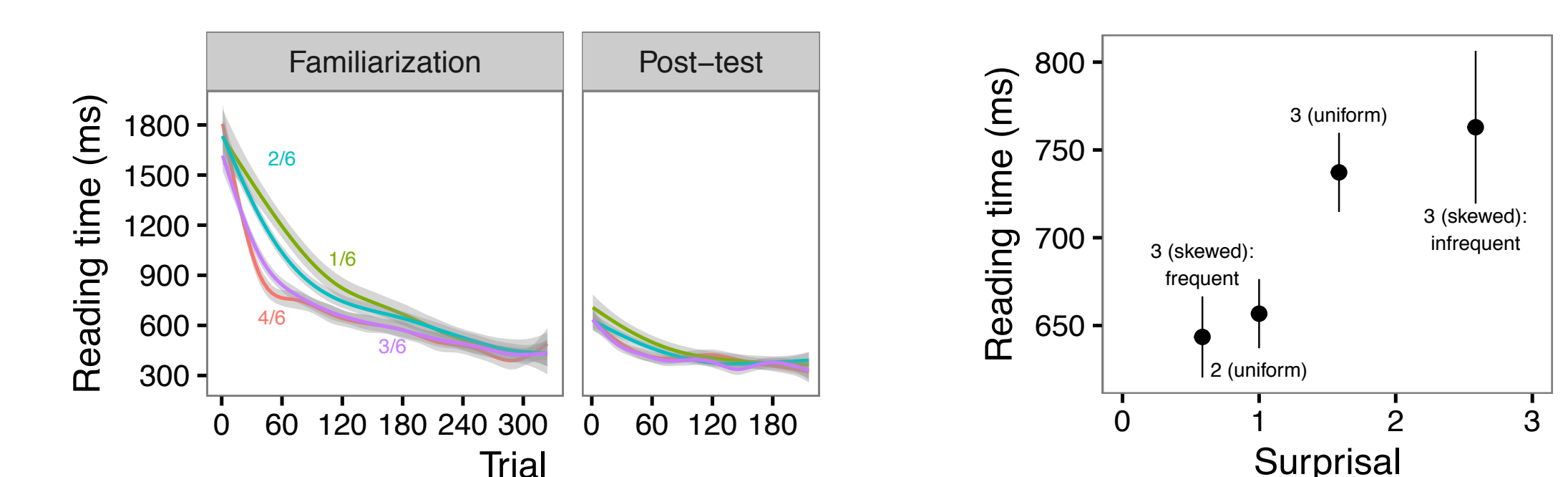
- Control for number of continuations
- Entropy gap smaller; no reduction comparison



Predictive glyph

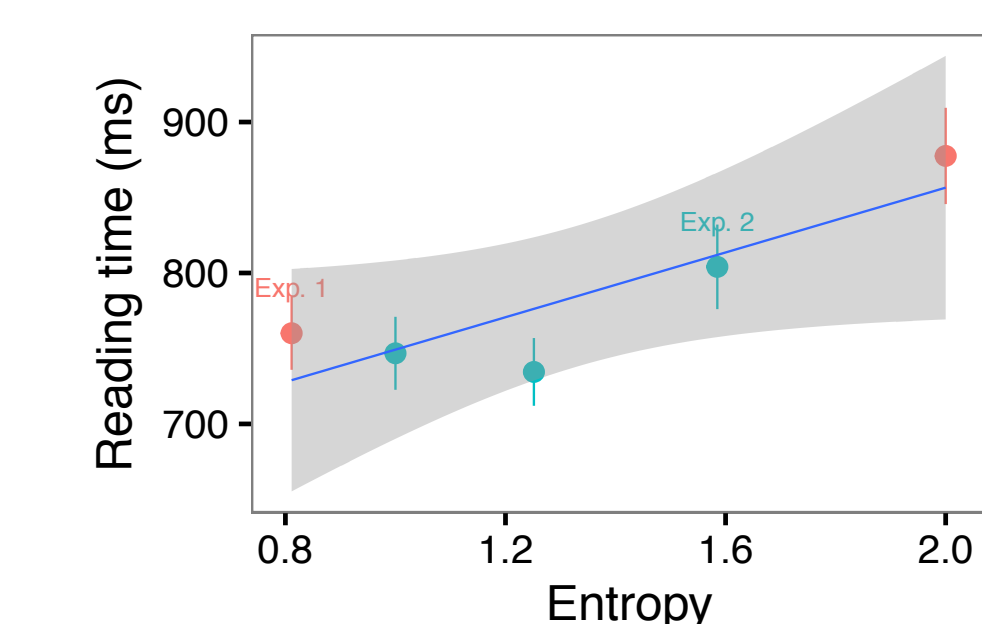


Predicted glyph

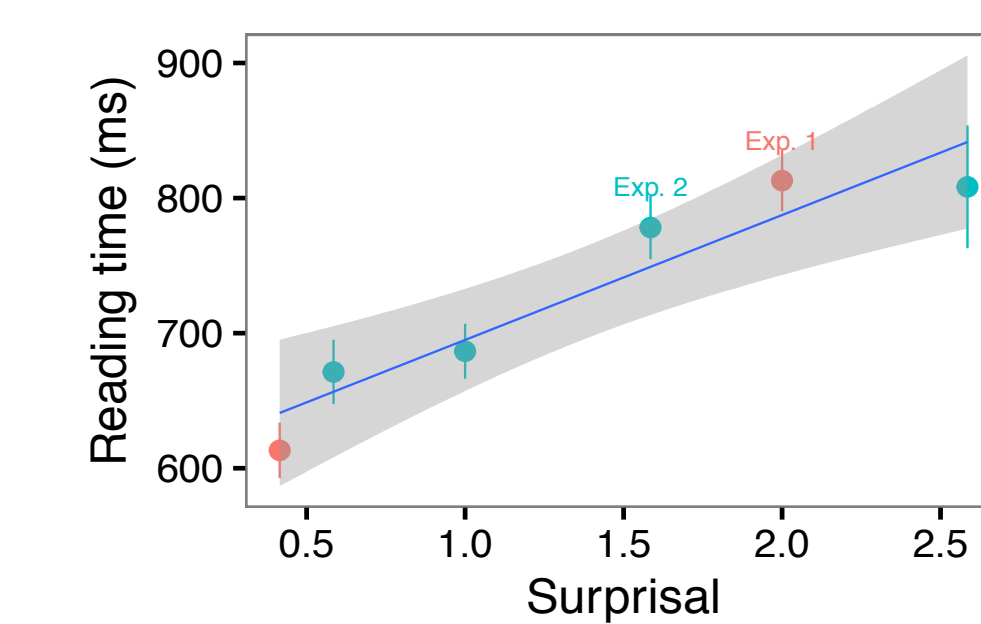


Combined analysis

Predictive glyph



Predicted glyph



Conclusions

- Graded, logarithmic predictability effects
- Higher entropy over predictions generally leads to longer reading times, but the effect may not be linear and/or depends on experimental context (what other predictive dependencies exist in the language)
- No evidence for an entropy reduction effect in this paradigm
- More empirical data is needed!