

Minimizing expected uncertainty in visual word recognition

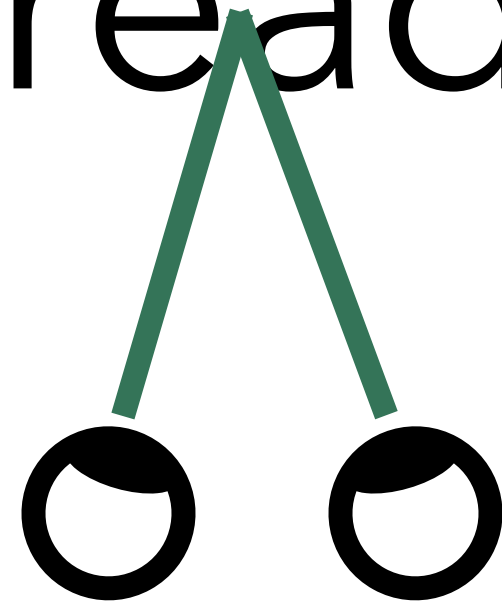
Are readers sensitive to the distribution of information across word forms?

Jon W. Carr & Davide Crepaldi

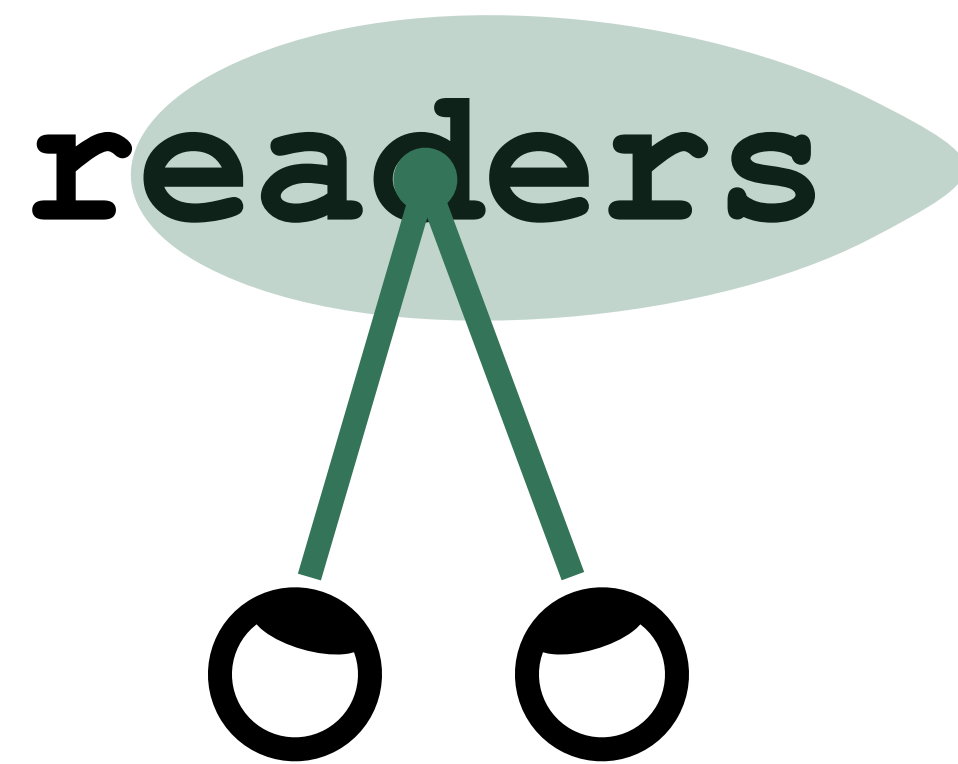
International School for Advanced Studies, Trieste, Italy



readers

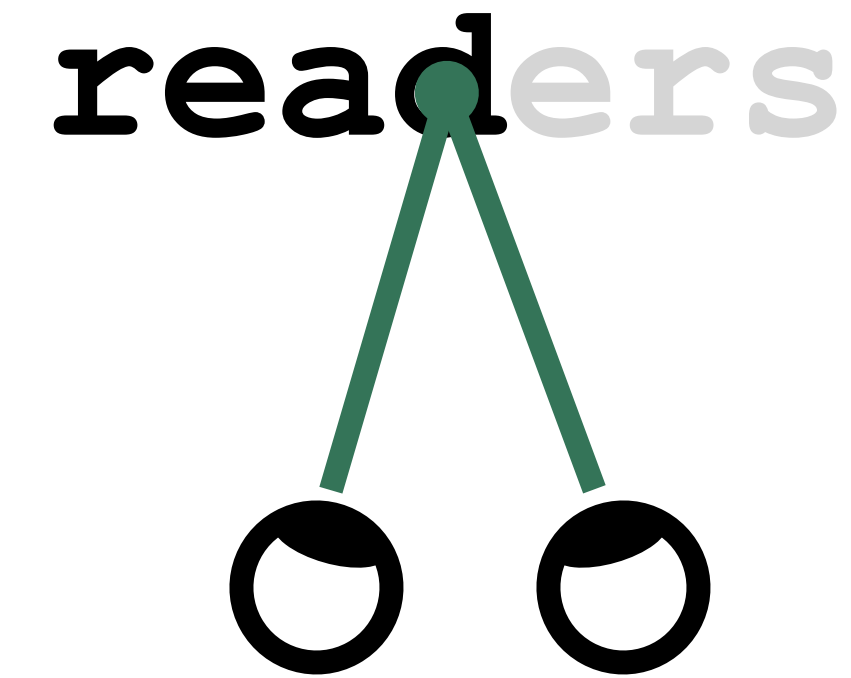


Perceptual account



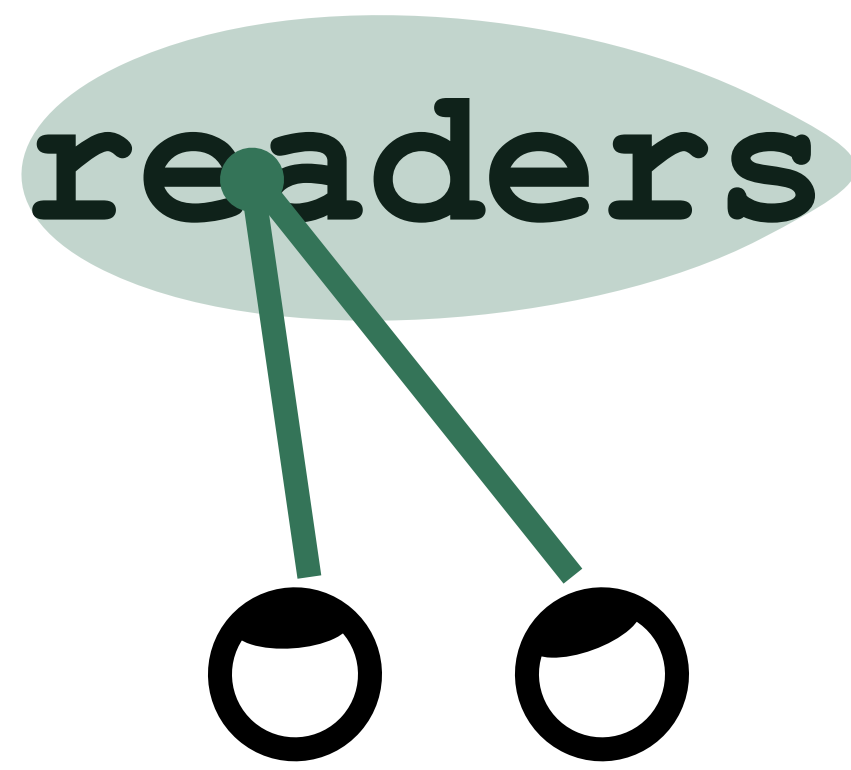
Visual span is asymmetric (right-visual-field advantage), so fixating left-of-center maximizes how much of the word is in view

Informational account



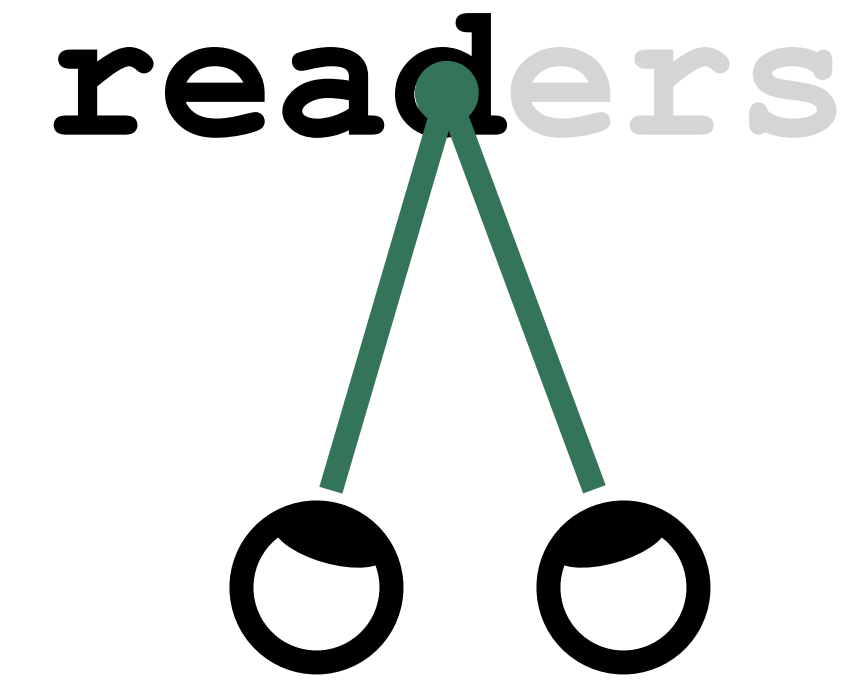
Words are typically more informative at the beginning, so fixating left-of-center places greater constraint on the possible words

Perceptual account



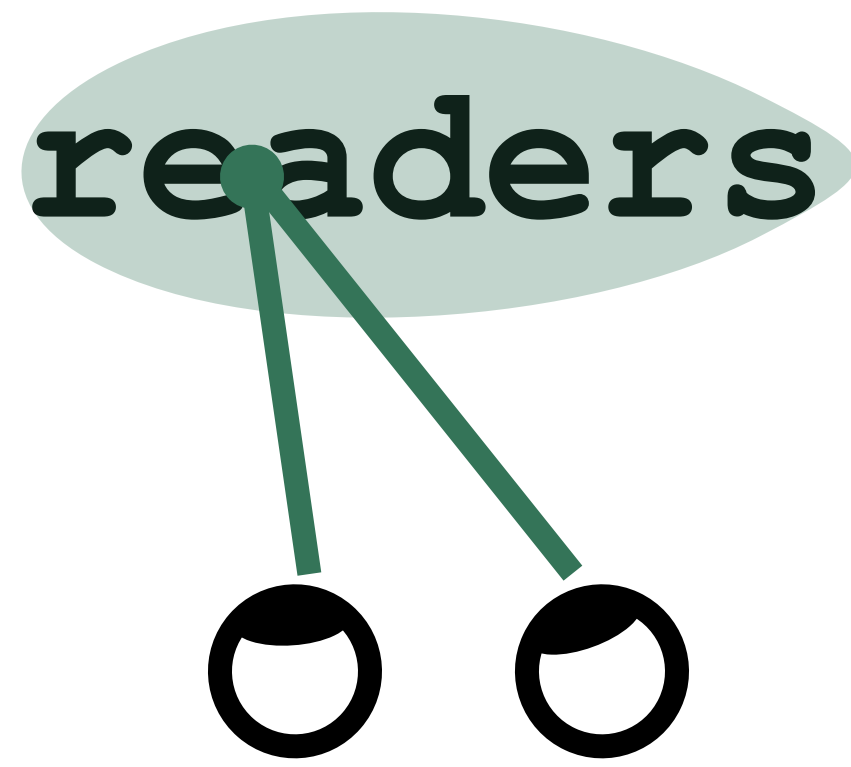
Visual span is asymmetric (right-visual-field advantage), so fixating left-of-center maximizes how much of the word is in view

Informational account



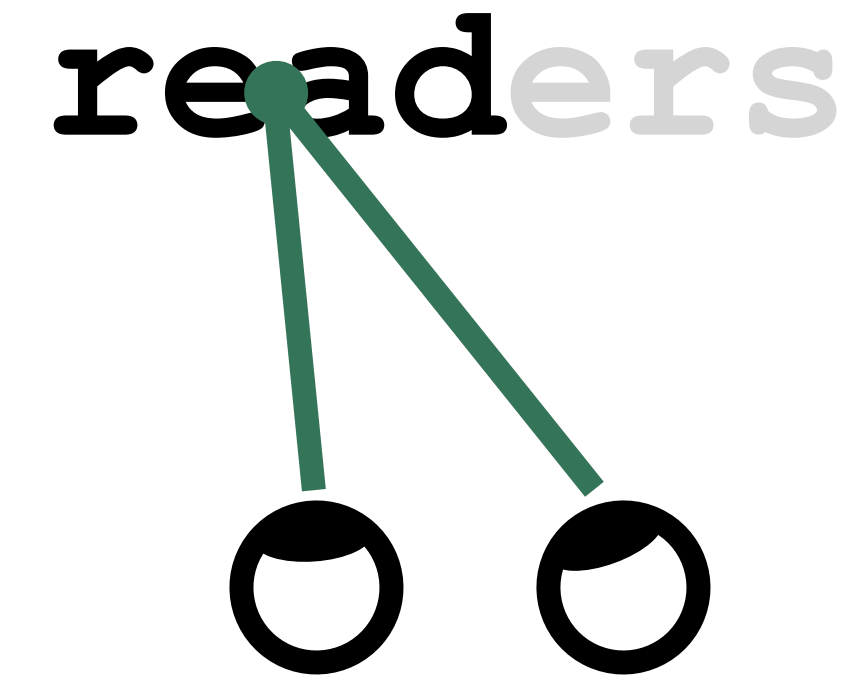
Words are typically more informative at the beginning, so fixating left-of-center places greater constraint on the possible words

Perceptual account



Visual span is asymmetric (right-visual-field advantage), so fixating left-of-center maximizes how much of the word is in view

Informational account



Words are typically more informative at the beginning, so fixating left-of-center places greater constraint on the possible words

*Are readers sensitive to how the
lexicon distributes information
across word forms?*



האוניברסיטה העברית בירושלים
THE HEBREW UNIVERSITY OF JERUSALEM

1. Dept. of Psychology, University of Toronto

Talia K. Shafir¹, Noam Siegelman², Ram Frost^{2,3,4}, Blair C. Armstrong^{1,4}

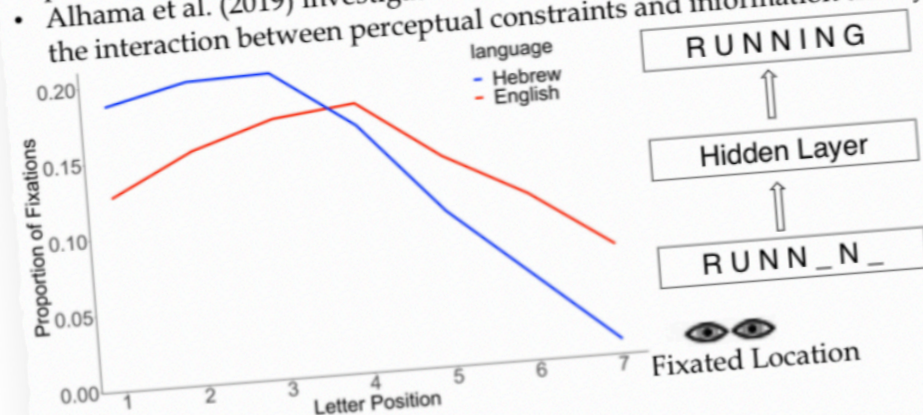
2. Haskins Laboratories - 3. Dept. of Psychology, The Hebrew University of Jerusalem - 4. BCBL

Haskins Laboratories



Introduction

- Considerable variability in the proportion of fixations across letter positions within and across languages during reading
- Alhama et al. (2019) investigated whether this variability was due to the interaction between perceptual constraints and information theory



Key Questions

- To what extent can the interaction of information theory and perceptual constraints explain the distributions of fixations across letter positions and languages?
- Can the account be generalized to predict the distribution of fixations across positions in other languages?

Big Picture Aim

- To generate a more universal account of eye-fixations during reading, thus reducing Anglo-centrism and potential model overfitting

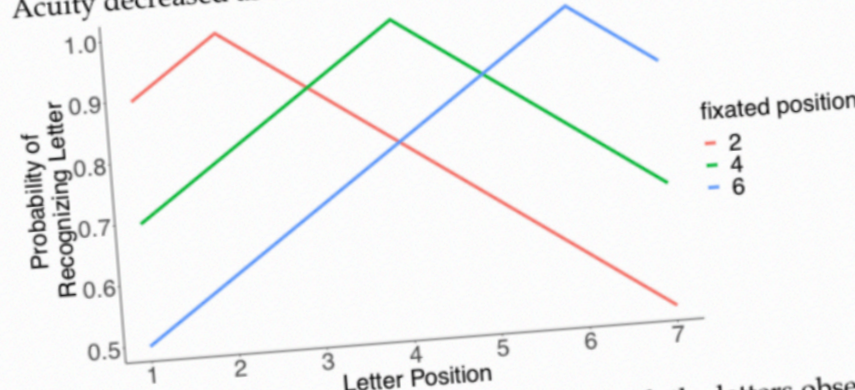
Methods

Target Stimuli

- The 3250 most frequent 7-letter words in 43-languages in Wikipedia from van Paridon & Thompson (2021)
- 7-letter words are sufficiently long to elicit differences in word recognition as a function of fixation location, but can still be perceived in a single fixation

Simulating Word Recognition at Different Fixation Locations

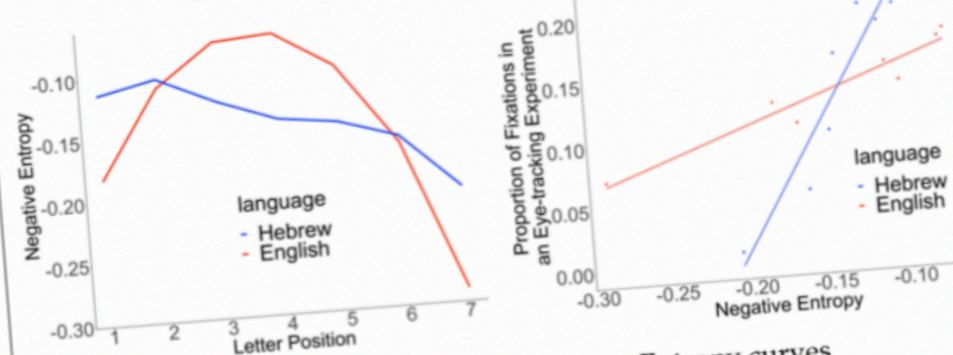
- Acuity decreased as a function of eccentricity from fixated location



- We identified all words that were consistent with the letters observed when fixating at each fixation location (40x)
- We computed the frequency-weighted likelihood of detecting each of these words and used these values to produce the Negative (normalized) Entropy over this distribution of probabilities as a rough proxy for fixation locations

Results

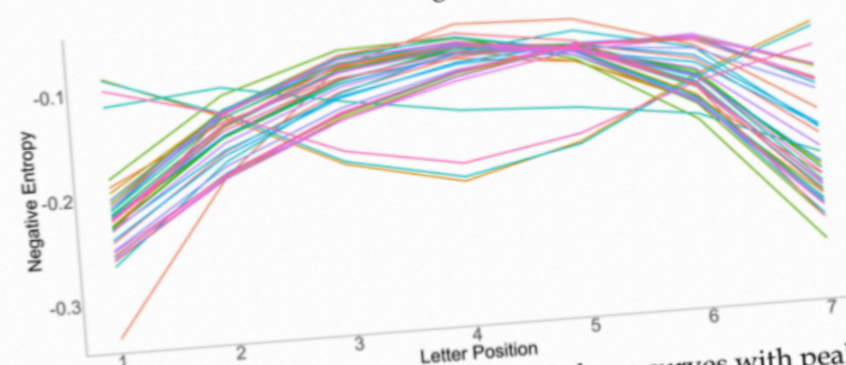
Relationship Between Fixations and Information Content in English and Hebrew



- English and Hebrew have distinct Negative Entropy curves
- Strong positive correlation between Negative Entropy and proportion of fixations in both languages ($r \geq .94$; $p < .0014$)

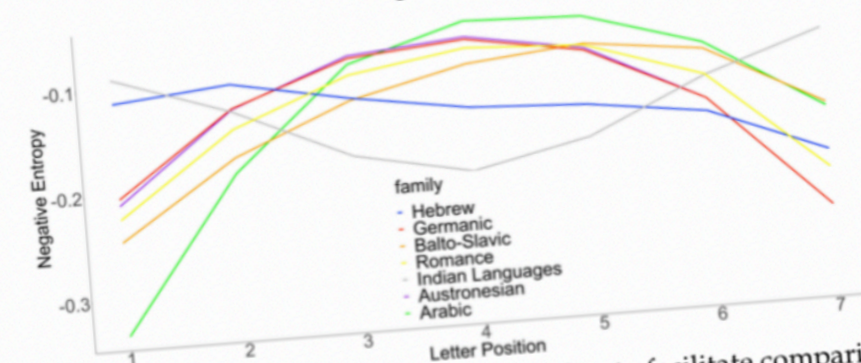
Predicting Distribution of Fixations across Letter Positions for 43 Languages

Negative Entropy as a Function of Letter Position by Language



- Most languages exhibit similar inverted U-shape curves with peak Negative Entropy slightly closer to word onset

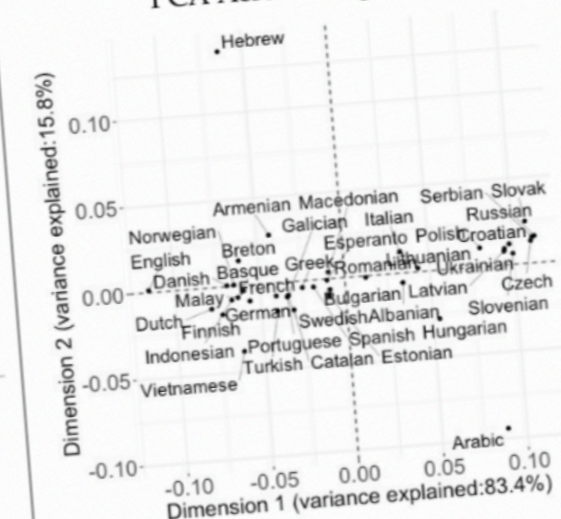
Negative Entropy as a Function of Letter Position by Language Family



- We took the average across language families to facilitate comparisons in the overall trends
- Main deviations from overall trend were:
 - Hebrew (not inverted U)
 - Arabic (few predicted fixations near onset)
 - Balto-Slavic languages (peak slightly closer to word offset)
 - Indian languages (syllabaries not captured by current approach)

Results

PCA Across Languages



- Two components reached significance (excluding Indian languages)
- With a few exceptions, the languages clustered based on their language families

Hierarchical Clustering Dendrogram



- Hebrew and Arabic are relative outliers in the space
- Slavic languages appear to group separately from other languages

Discussion

- Broad similarities in predicted fixation distributions across a diverse set of languages from 43 languages
- Success in capturing differences in reading behavior across languages and Hebrew behavior would be a fruitful if tested in a controlled experiment
- These results suggest that a model of reading behavior based on perceptual constraints and information theory can capture the overall trends we look for

This work was supported by the National Science Foundation (grant number 06310) to Blair Armstrong.

Alhama, R. G., Siegelman, N., & Thompson, S. A. (2019). Word recognition: A model of the interaction of perceptual constraints and information theory. *Psychological Review*, 126, 83-89.

van Paridon, J., & Thompson, S. A. (2021). Research Methods, 50.

Shafir et al.

P5-15-1849

Left-heavy lexicon

SNYBEVS
STOBEVS
SGUPEVS
SKAPEVS
SGYDIVS
SNODIVS
SKUMIVS
STAMIVS



High
information
content



Low
information
content

Right-heavy lexicon

SVEBYNS
SVEBOTS
SVEPUGS
SVEPAKS
SVIDYGS
SVIDONS
SVIMUKS
SVIMATS



Low
information
content



High
information
content

Left-heavy lexicon

SNYBEVS

STOBEVS

SGUPEVS

SKAPEVS

SGYDIVS

SNODIVS

SKUMIVS

STAMIVS



High
information
content



Low
information
content

Right-heavy lexicon

SVEBYNS

SVEBOTS

SVEPUGS

SVEPAKS

SVIDYGS

SVIDONS

SVIMUKS

SVIMATS



Low
information
content



High
information
content

Prediction

SXXXXXS

Left-heavy lexicon

SNYBEVS

STOBEVS

SGUPEVS

SKAPEVS

SGYDIVS

SNODIVS

SKUMIVS

STAMIVS

High
information
content

Low
information
content

Right-heavy lexicon

SVEBYNS

SVEBOTS

SVEPUGS

SVEPAKS

SVIDYGS

SVIDONS

SVIMUKS

SVIMATS

Low
information
content

High
information
content

Prediction

SX●XXX●XS

Left-heavy lexicon

SNYBEVS

STOBEVS

SGUPEVS

SKAPEVS

SGYDIVS

SNODIVS

SKUMIVS

STAMIVS

High
information
content

Low
information
content

Right-heavy lexicon

SVEBYNS

SVEBOTS

SVEPUGS

SVEPAKS

SVIDYGS

SVIDONS

SVIMUKS

SVIMATS

Low
information
content

High
information
content

Prediction

SX●X●XXS

Left-heavy lexicon

SNYBEVS

STOBEVS

SGUPEVS

SKAPEVS

SGYDIVS

SNODIVS

SKUMIVS

STAMIVS

High
information
content

Low
information
content

Right-heavy lexicon

SVEBYNS

SVEBOTS

SVEPUGS

SVEPAKS

SVIDYGS

SVIDONS

SVIMUKS

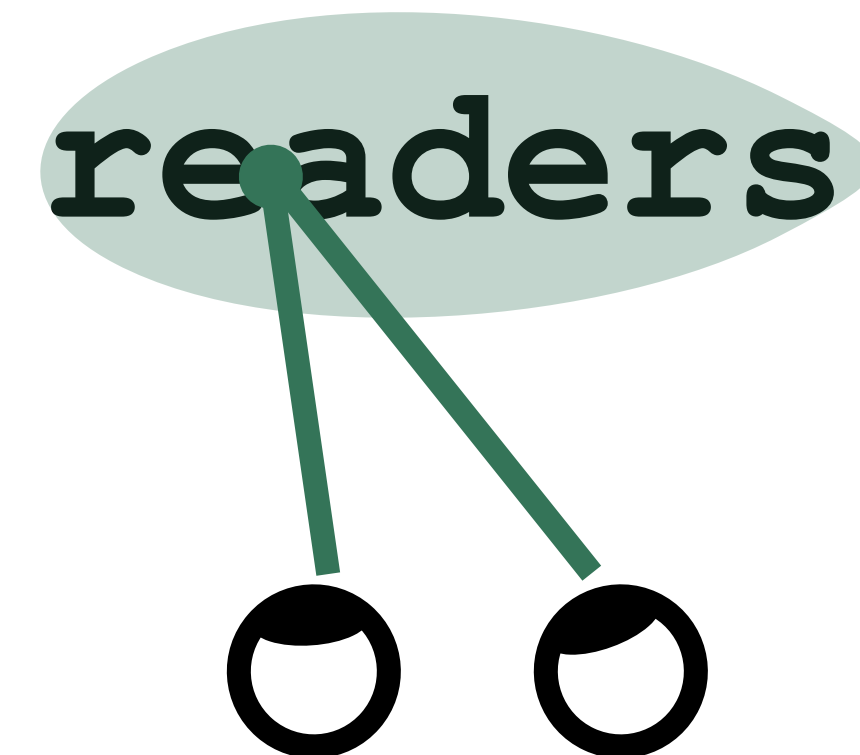
SVIMATS

Low
information
content

High
information
content

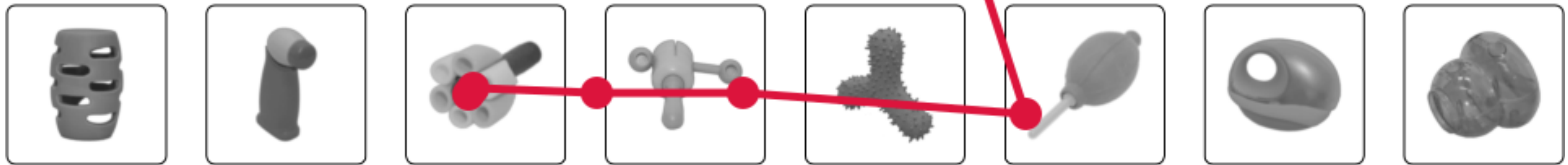
Prediction

SX●X●XXS

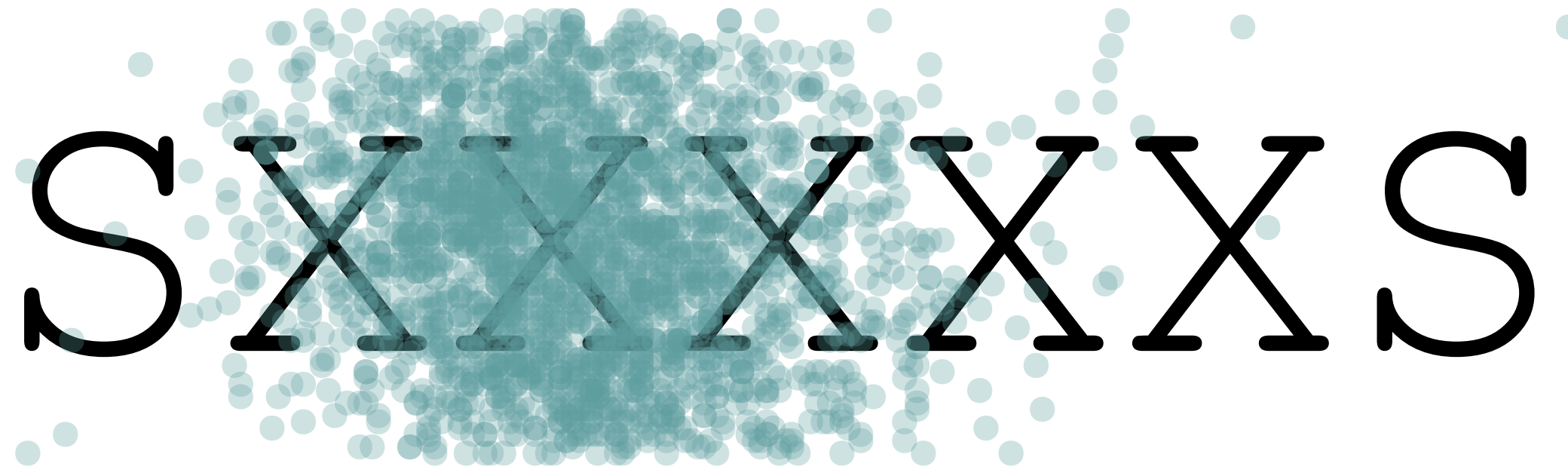


initial landing position

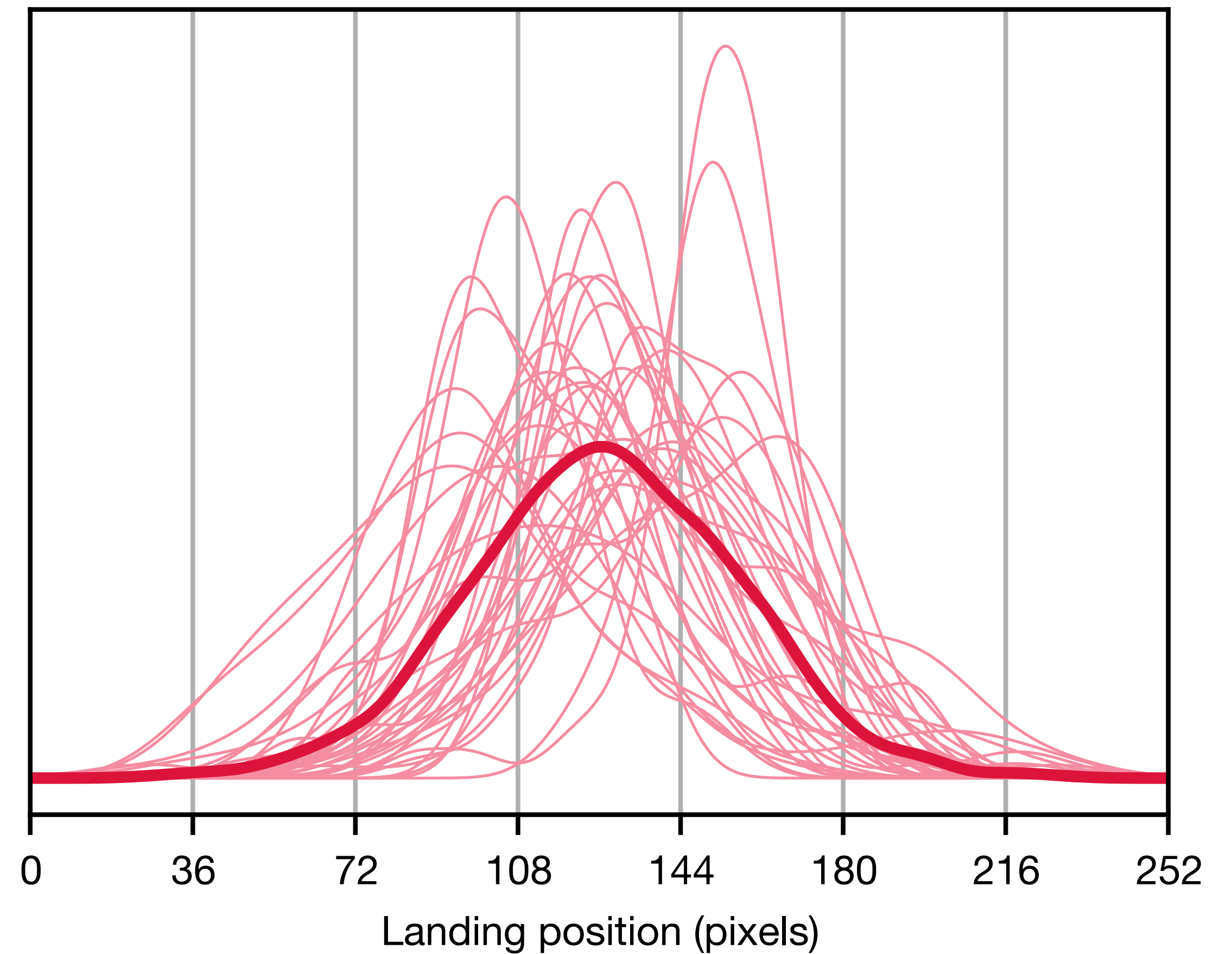
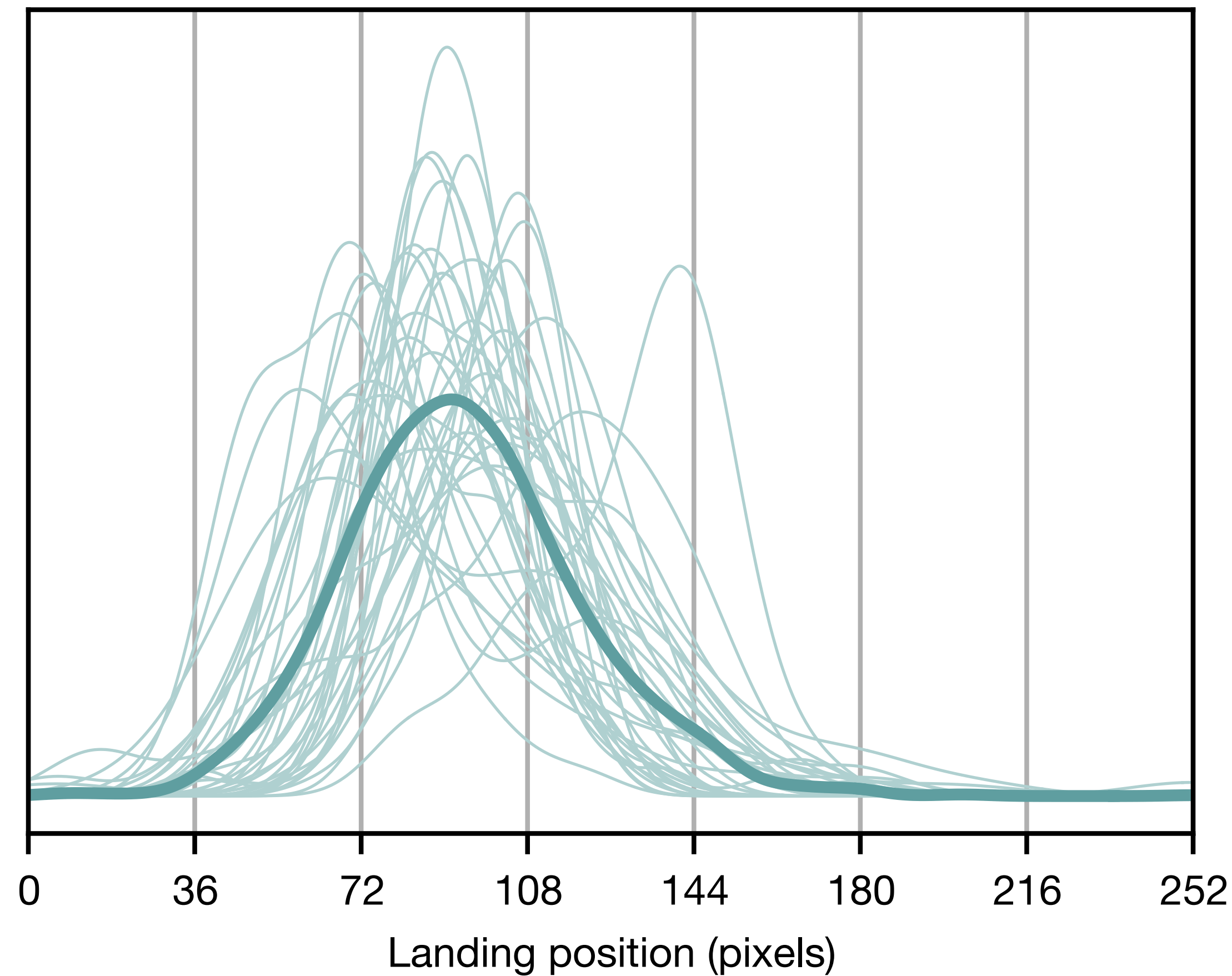
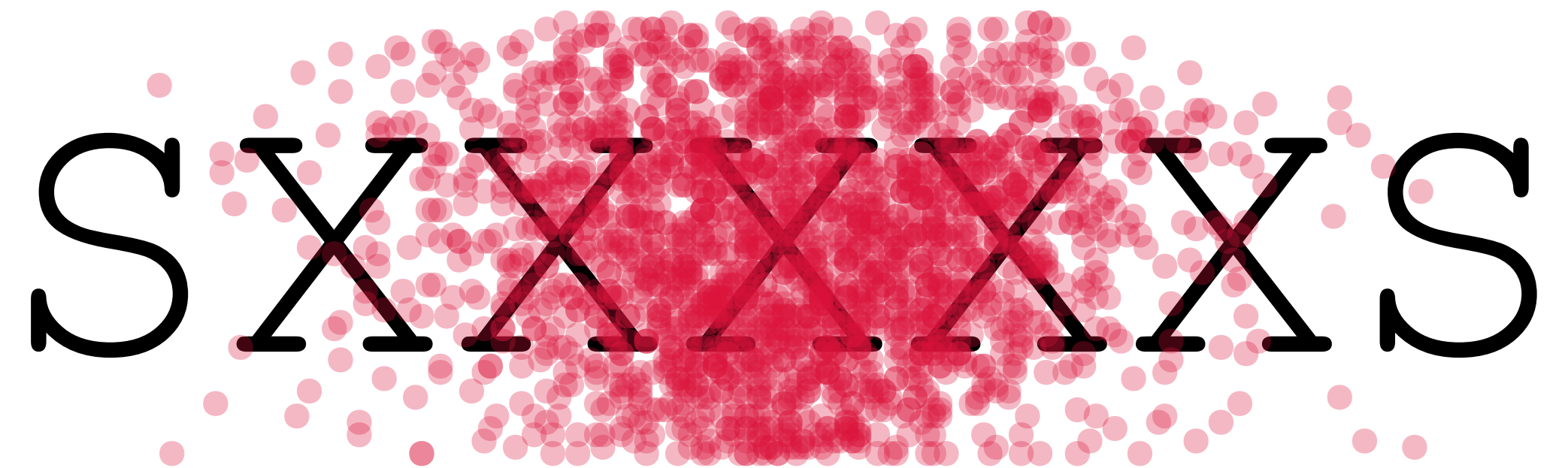
SMADEPS



Participants exposed to the left-heavy lexicon



Participants exposed to the right-heavy lexicon



*Are readers sensitive to how the
lexicon distributes information
across word forms?*

