

Lexical Retrieval and Probabilistic Inference Dissociate in Naturalistic Reading

Cory Shain, MIT

BACKGROUND

The processing difficulty of words is related to their **frequency** and **predictability** in context. **But why?**

Two-Process View: One process retrieves items from the mental lexicon and another preactivates items from context (Staub 2015).

Prediction: Dissociation between frequency and predictability effects.

One-Process View: Comprehension difficulty is driven by surprisal (Hale 2001, Levy 2008), of which frequency effects are a subtype.

Prediction: No dissociation between frequency and predictability effects.

Empirical Record:

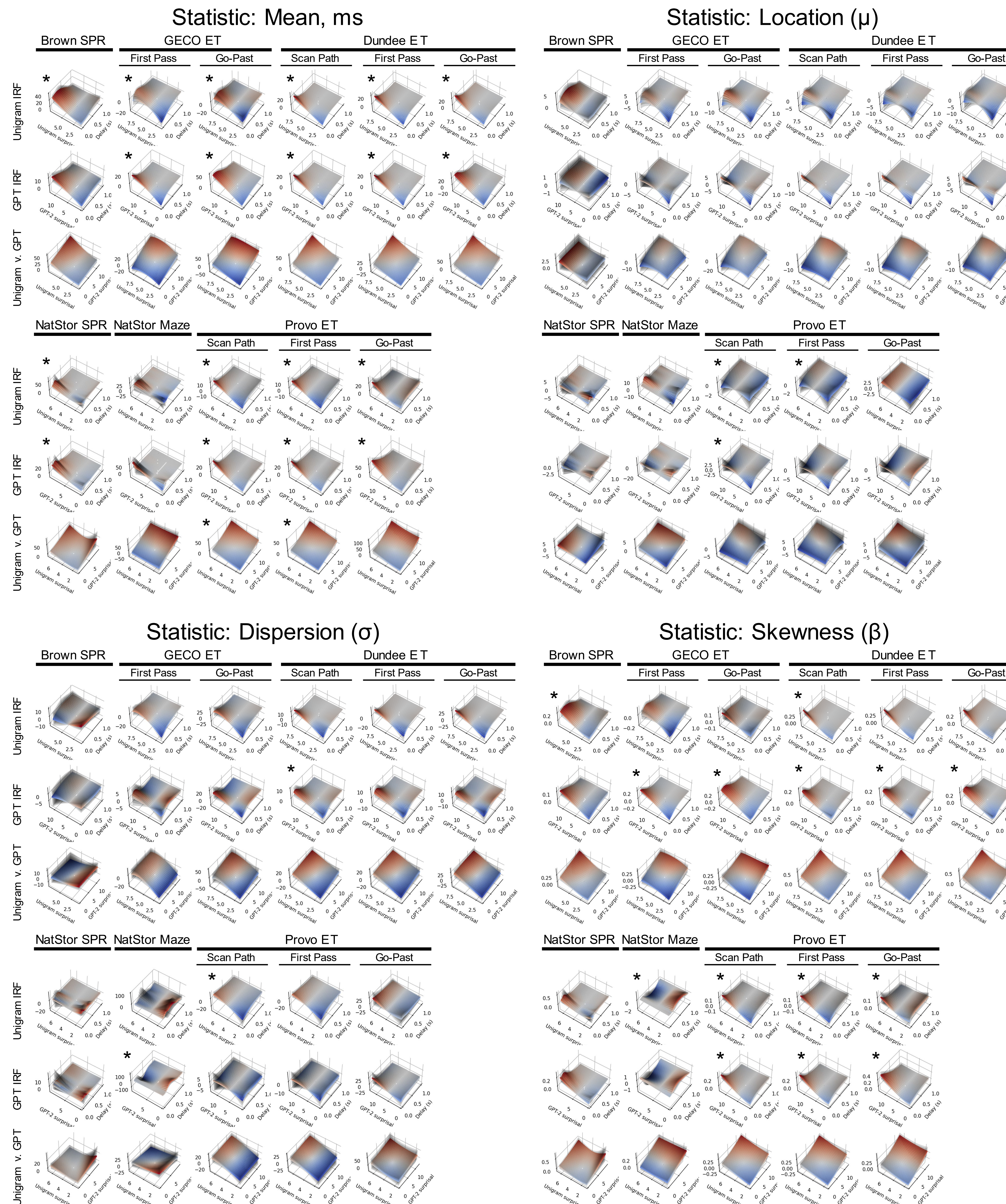
Controlled studies: Generally favor two-process view (Staub 15)

Naturalistic studies: Mixed results (Shain 19 vs. Goodkind & Bicknell 22)

Distributional effects: Some evidence that frequency affects both location and skewness while predictability only affects skewness (Staub et al 10; Staub 11). Never directly tested.

MAIN QUESTION

Do frequency and predictability dissociate in naturalistic reading?



METHOD

6 public datasets (Brown, Dundee, GECO, Natural Stories, Natural Stories Maze, Provo)

3 modalities (eye-tracking, self-paced reading, Maze)

>400 participants

>2M fixations

Flexible nonlinear regression (CDRNNs; Shain & Schuler 22)

Ensemble size: 10

ExGaussian predictive distribution

Out-of-sample statistical comparisons

RESULTS

Dissociable, additive (non-interacting) frequency and predictability effects

Little evidence of distributional differences between frequency and predictability (c.f. Staub et al 10, Staub 11)

CONCLUSION

Frequency and predictability effects dissociate in naturalistic reading

Supports either:

1. **Two-process** view
2. Modified **one-process** view to allow lossy memory (e.g., Hahn et al 22)