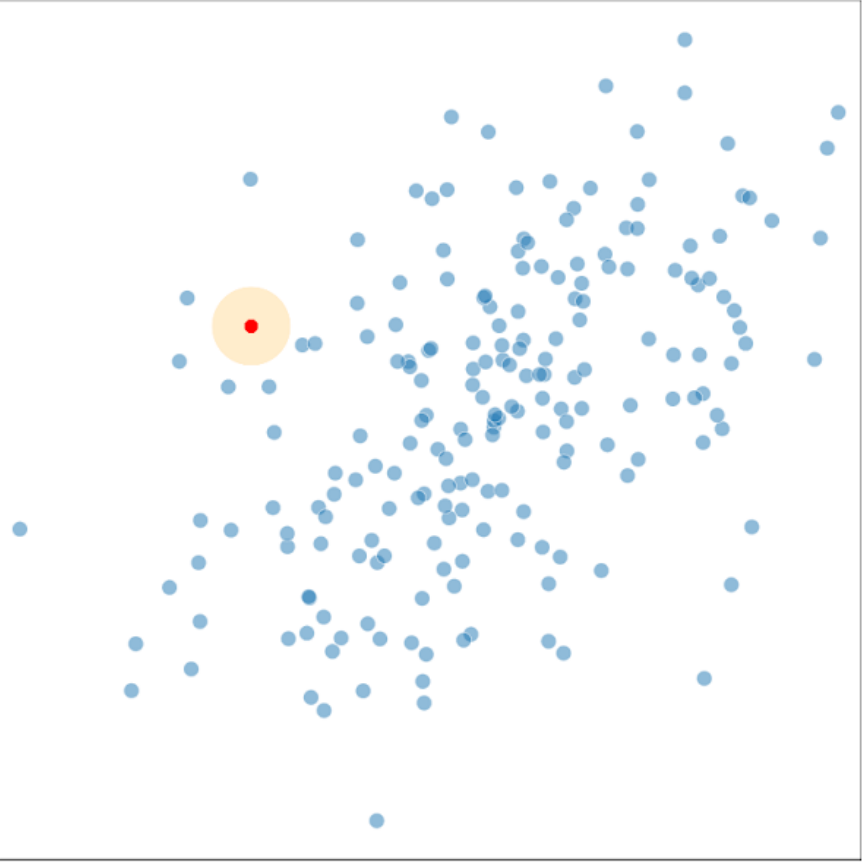


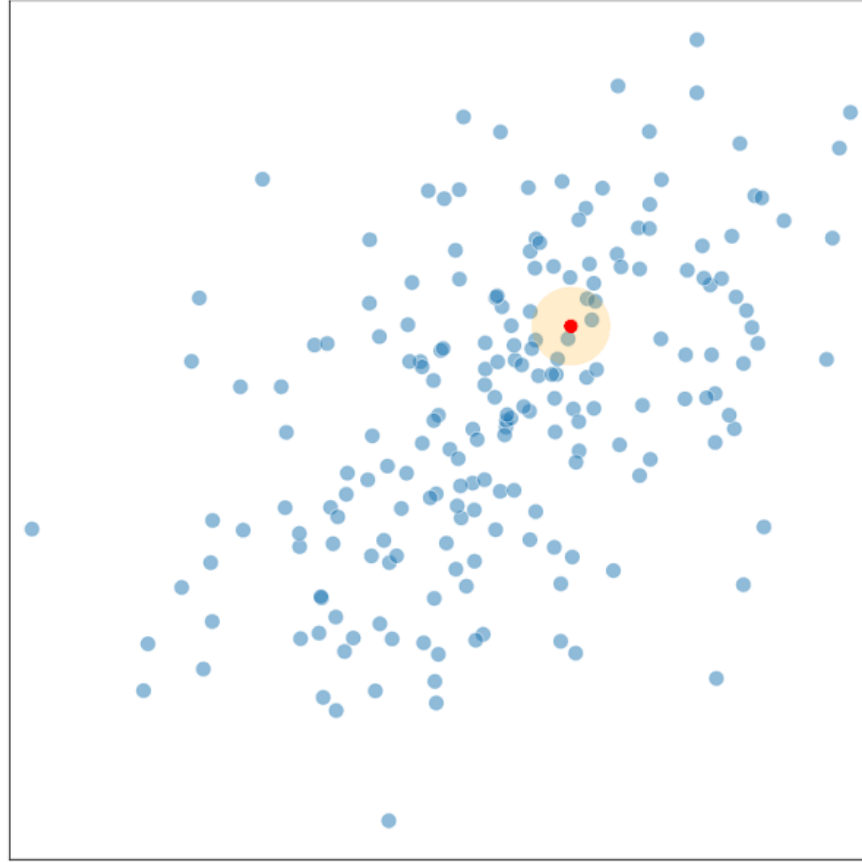
Background: Noisy Memory Representations

- Memory representations **accumulate noise** over time [1]
- Familiarity judgments may be **probabilistic**, taking into account uncertainty [1]
- Distinctive** items are more likely to be remembered [2]
- Language Models** provide a continuous representation of sentence meaning
- Prediction:** sentences in dense parts of LMs' meaning space are less likely to trigger false alarms on first presentation and more likely to be recognized subsequently

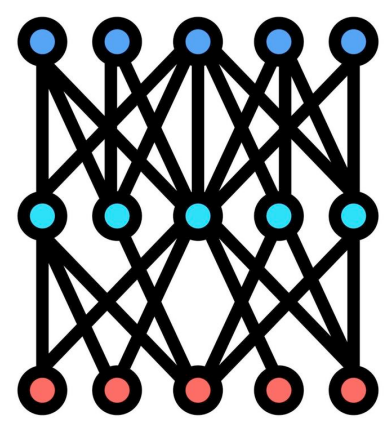
A Sparse Part of the Meaning Space



B Dense Part of the Meaning Space



Sentence



Model Representation

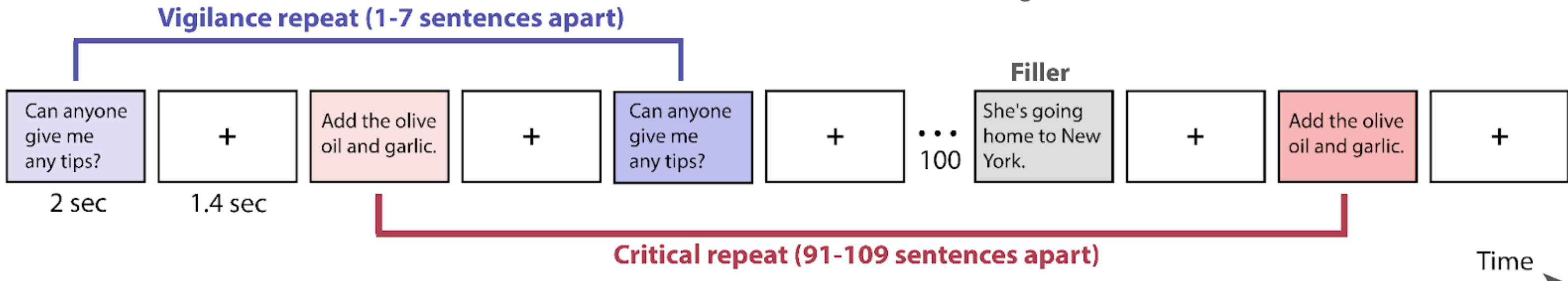
We operationalize **semantic distinctiveness** as: average cosine distance in model representation space to large corpus of sentences

For illustrative purposes

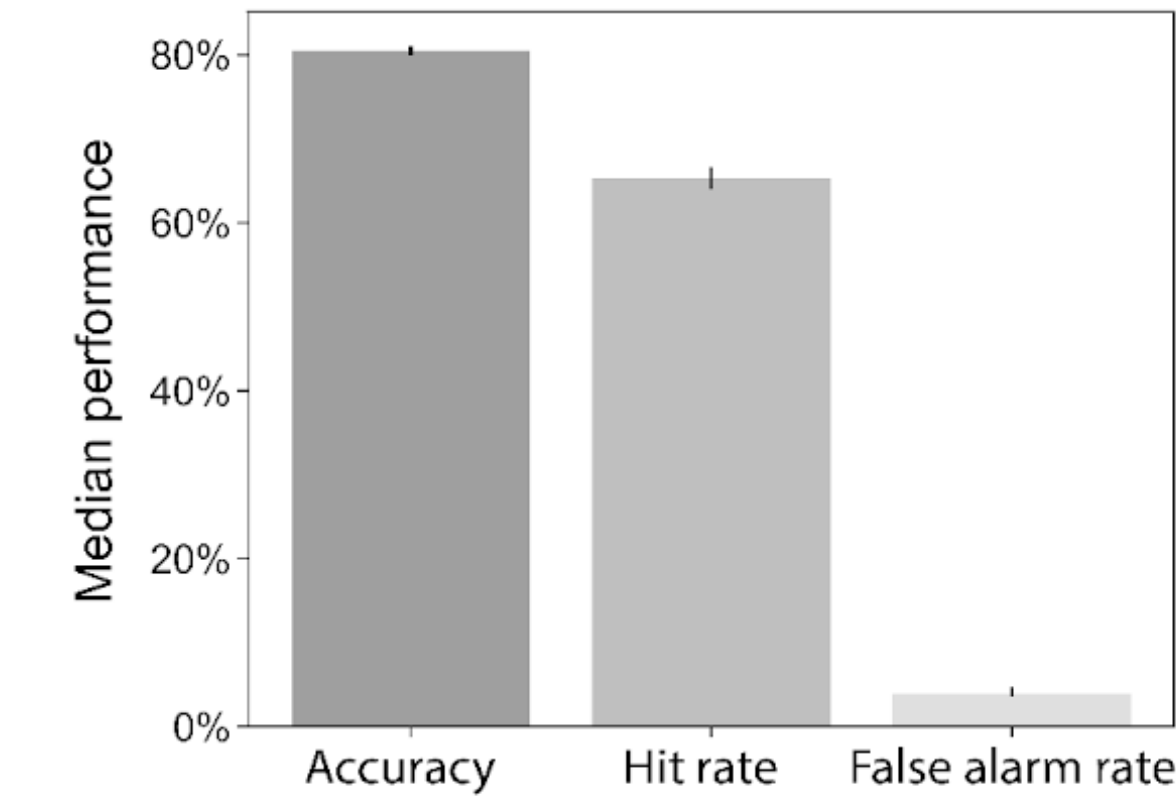
Behavioral Memory Experiment

N=500 Prolific participants
2500 target sentences + 1000 fillers

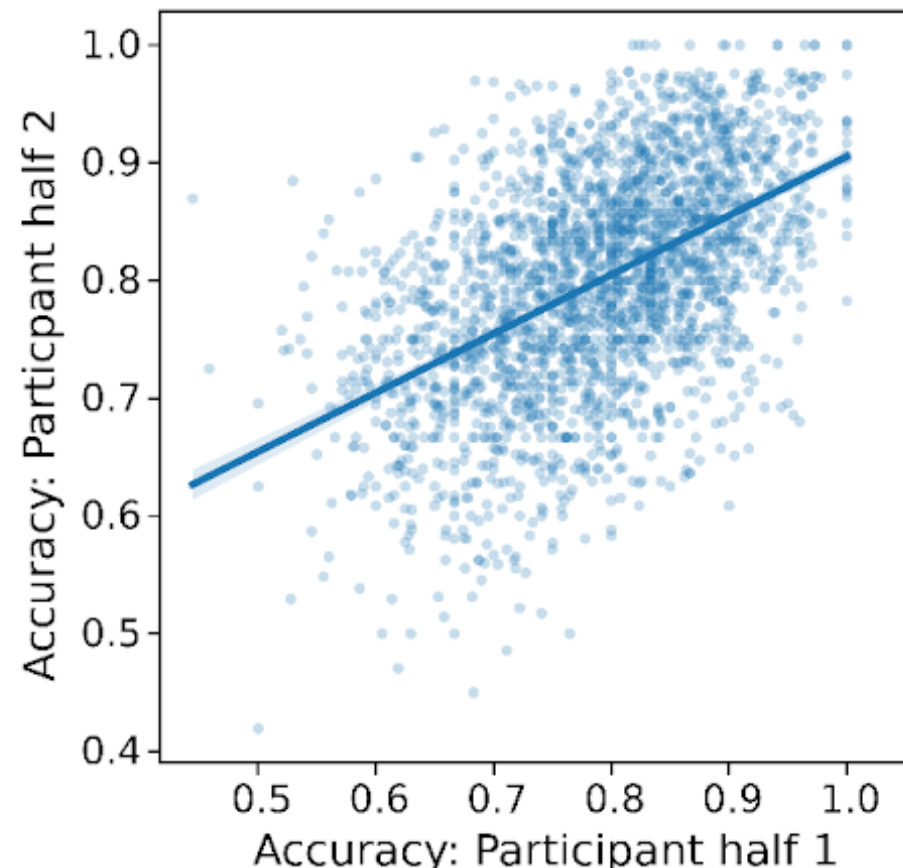
A Schematic of Experimental Paradigm



B Memorability of Sentences



C Consistency of Sentence Memorability Scores Across Participants



Each dot is one sentence: accuracy from half of participants predicts accuracy from other half

D Most and Least Memorable Sentences

Most memorable

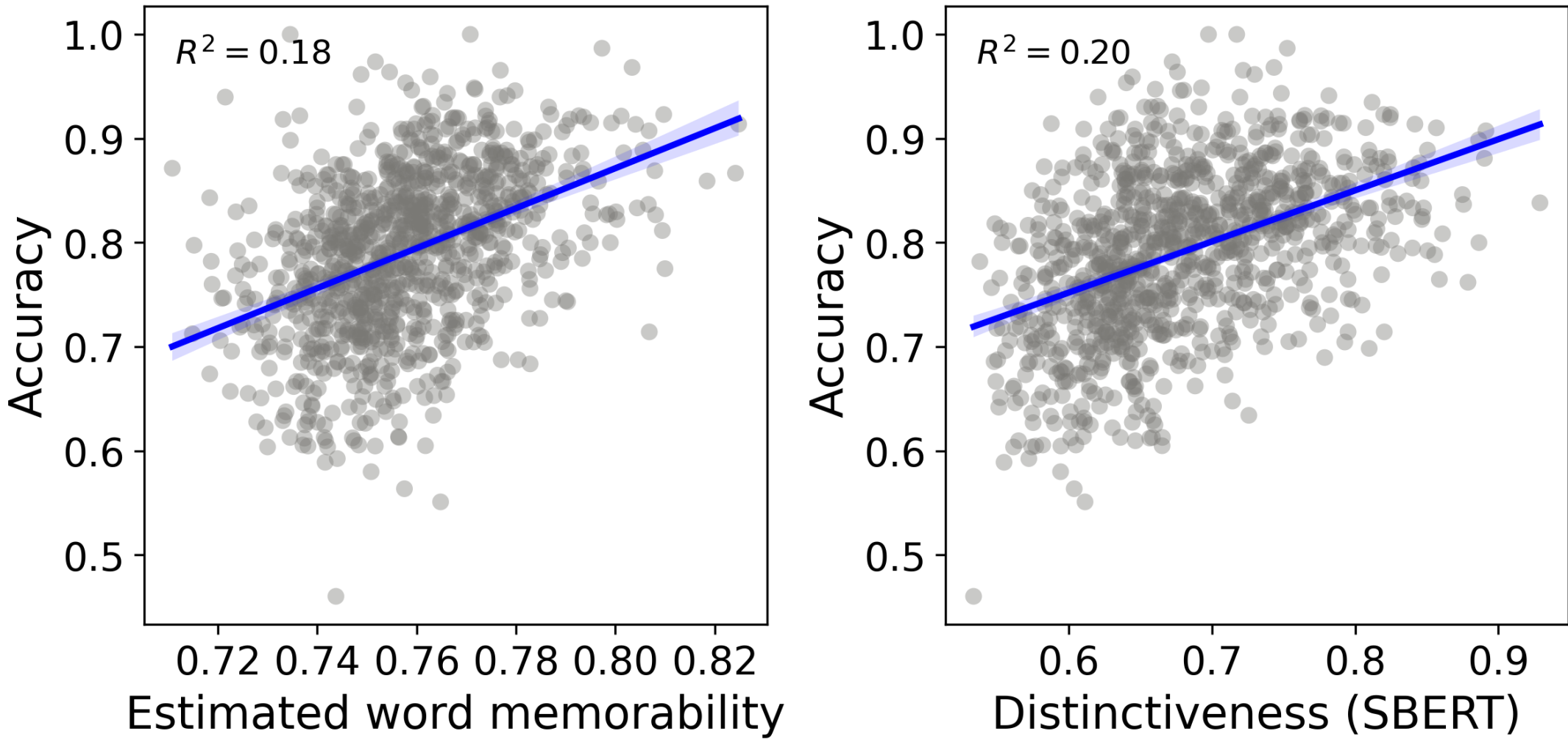
Wait a sec, Nicky, Nicky, Nicky.
Homer Simpson is hungry, very hungry.
Does olive oil work for tanning?
These mosquitos are -- well, guinea pigs.
Goddammit, where are you from, Mars?
Macbook Pro repair in Niagara region.
They stabbed her fifty three times!
Now let's talk about Donkey Kong.
Top with mozzarella, mushrooms, and peppers.
Every cloud has a blue lining!

Least memorable

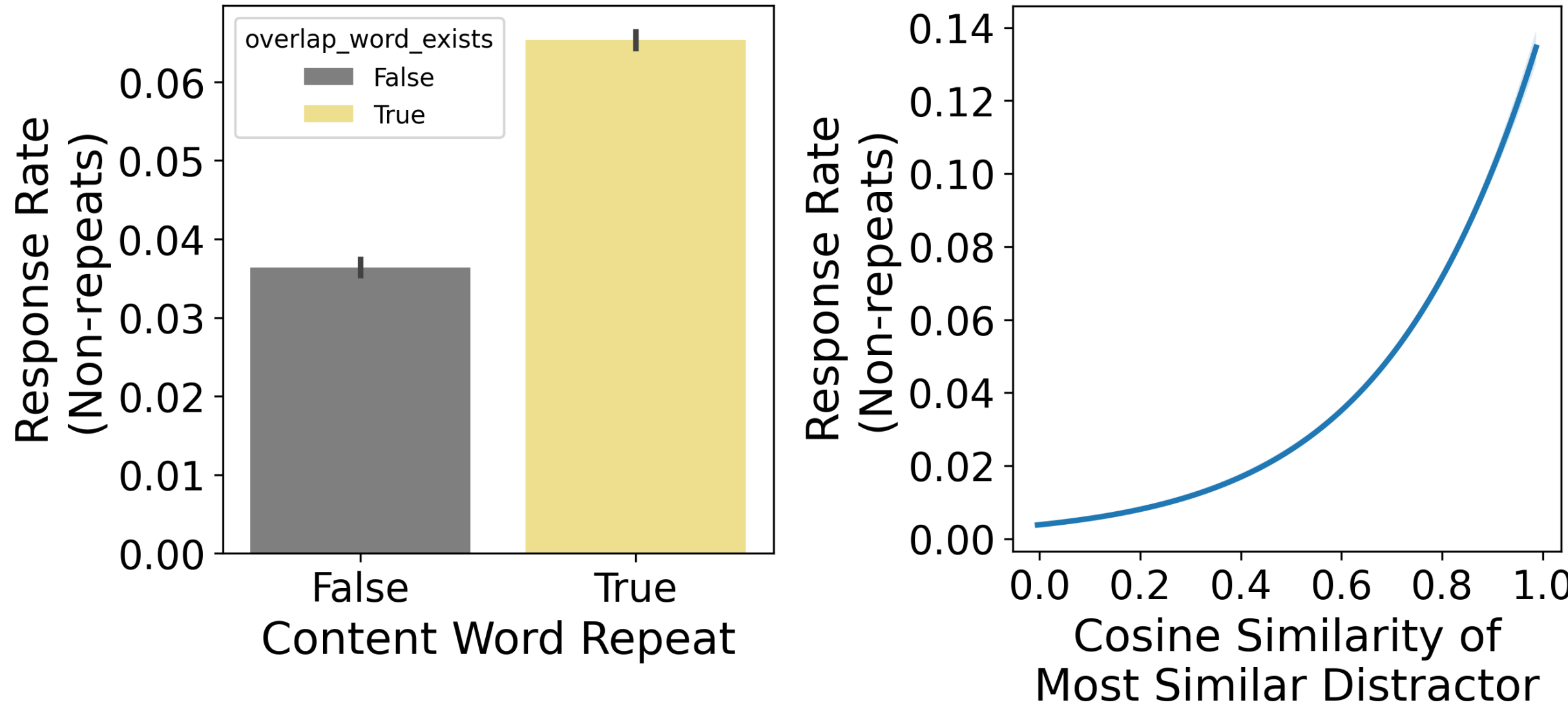
I just couldn't leave without it.
We want to make it better.
But still it wasn't nearly enough.
It must be getting more serious
But it made me feel better..
Therefore to hold meetings by them.
Another manner is to time entries.
But give it to me again.
You still had to prove yourself.
Unless he'd been afraid they'd know.

Results

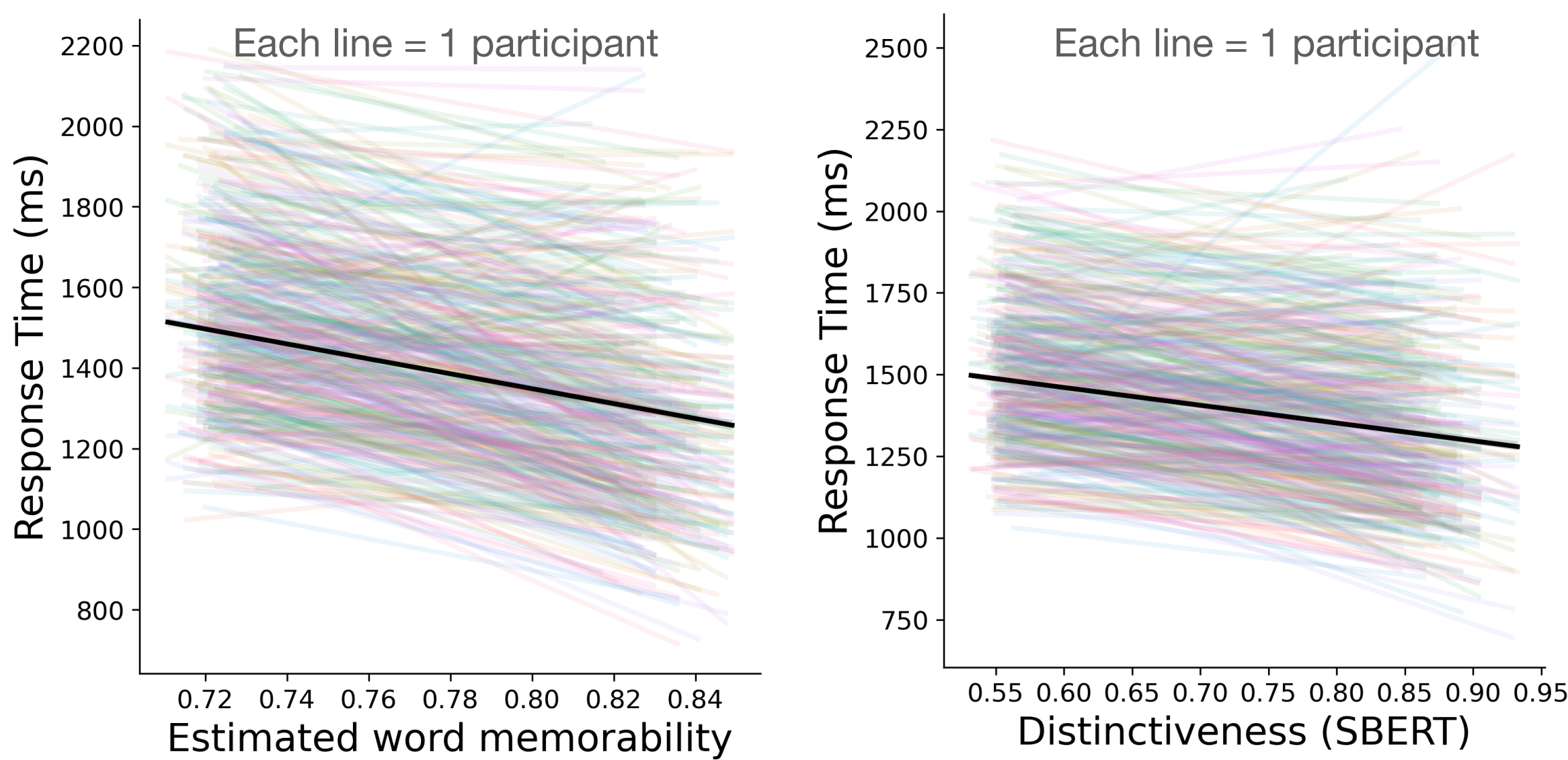
Mean Accuracy



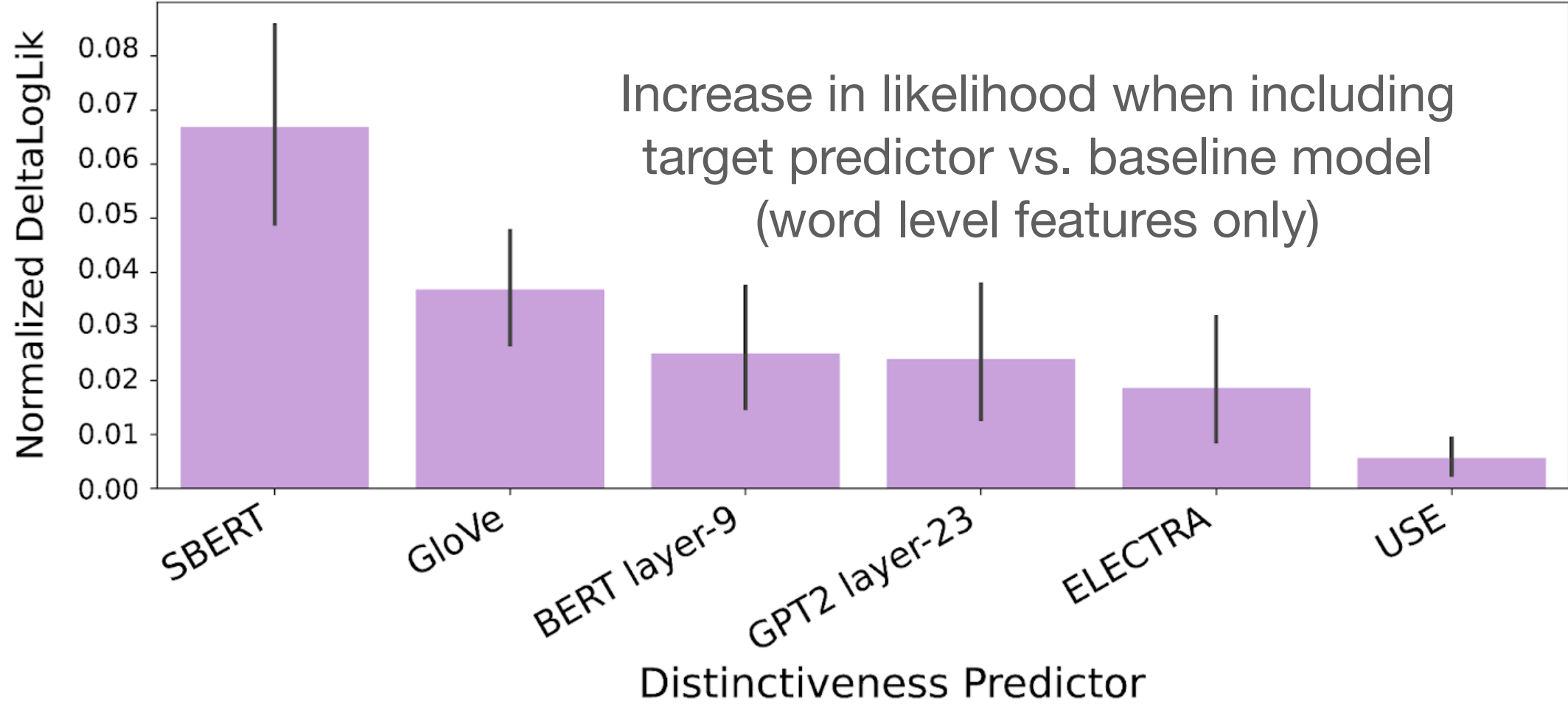
False Alarms



Response Time



Delta Log-Likelihood Predictor Performance of Sentence Memorability



Note: Consistent with the above visualizations, distinctiveness was a significant predictor of memory accuracy and speed when controlling for word-level features, using linear regression for item-level accuracy, linear mixed effects models for trial-level reaction time and logistic mixed-effects models for trial-level false alarms.

Predictors of Memorability

Predictor	Description	Top Sentence	Bottom Sentence
Word Memorability	Estimated using data from [3]	Herbed chevre, pomodoro sauce, basil toast.	Not just physically but intellectually, emotionally.
SBERT [4]	Sentence embedding model with strong semantic performance	Texas steak dinner on a stick?	You've managed to edit a value!
GloVe [5]	Sentence representation that only considers word identities	Tantalum, Titanium, Hastelloy, Inconel and Monel.	You have to make it last.
BERT [6]	Hidden Layer 9 Activations	A scream spilled from Kylie's lips.	And getting better all the time.
GPT-2 [7]	Hidden Layer 23 Activations	I think it's overblown,' he said.	Nothing you would find very remarkable.
ELECTRA [8]	Sentence embedding model	One, two, three, four, five, six-	But those plastic ties are unbreakable.
USE [9]	Universal Sentence Encoder: Sentence embedding model	Implement continuous and routine improvement ideas.	He sneered contemptuously at the Daimons.

Conclusions

- Distinctiveness** — a single scalar predictor computed using **sentence representations from LMs** — predicts empirical recognition memorability, even when controlling for word-level memorability
- Distinctiveness w.r.t. both **general language statistics** and **experimental context** predicts performance; representation uncertainty may explain memory performance across timescales