Experimental versus In-Corpus Variation in Referring Expression Choice

T. Mark Ellison & Fahime Same (2024)

SFB 1252
PROMINENCE
IN LANGUAGE

Universität
zu Köln

Referring Expressions

point at things
but can have
one of three different
referring expression forms (REFs)

- proper name: **Jinjin**
- definite description: **the small quokka**
- pronoun: **she**

Variation in Reference

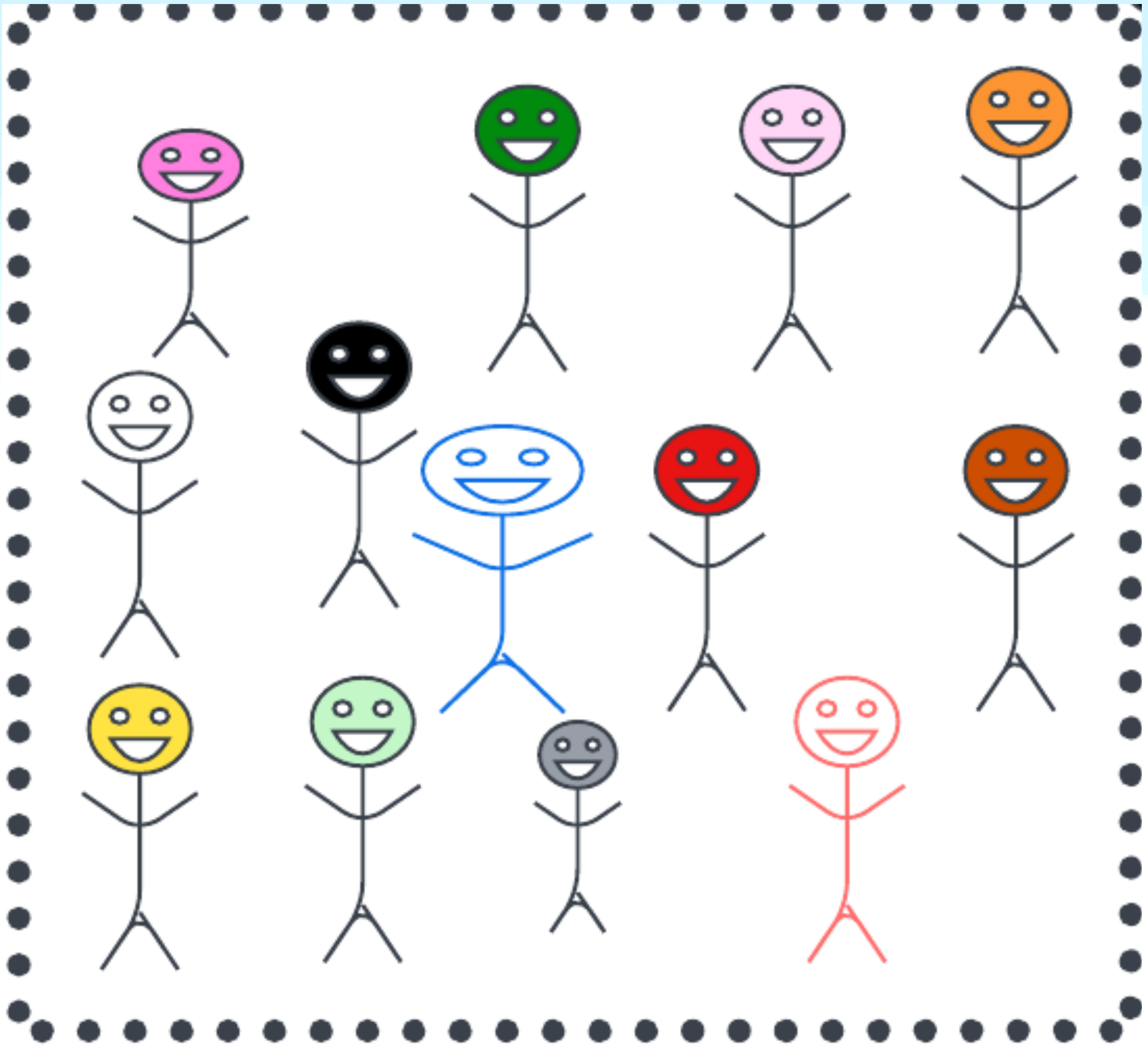
and counterfactuals

- What REF options does the author have at the moment of composition?
- Pronouns are sometimes too ambiguous
- Proper names and definite descriptions are sometimes too cumbersome
- But sometimes multiple options can fit?

Windows onto Variation

Experimentation

- Get some willing participants
- Show them the text with gaps where the referring expressions for the topic are
- Ask them to fill in those gaps
- See how much their choices vary
- Castro Ferreira et al. 2016



Homer_Simpson (born May 12 1956) is the main protagonist and one of the five main characters of The Simpsons series (or show). **Homer_Simpson** is the spouse of Marge Simpson and father of Bart, Lisa and Maggie Simpson. **Homer_Simpson** is overweight (said to be ~240 pounds), lazy, and often ignorant to the world around **Homer_Simpson**.

→

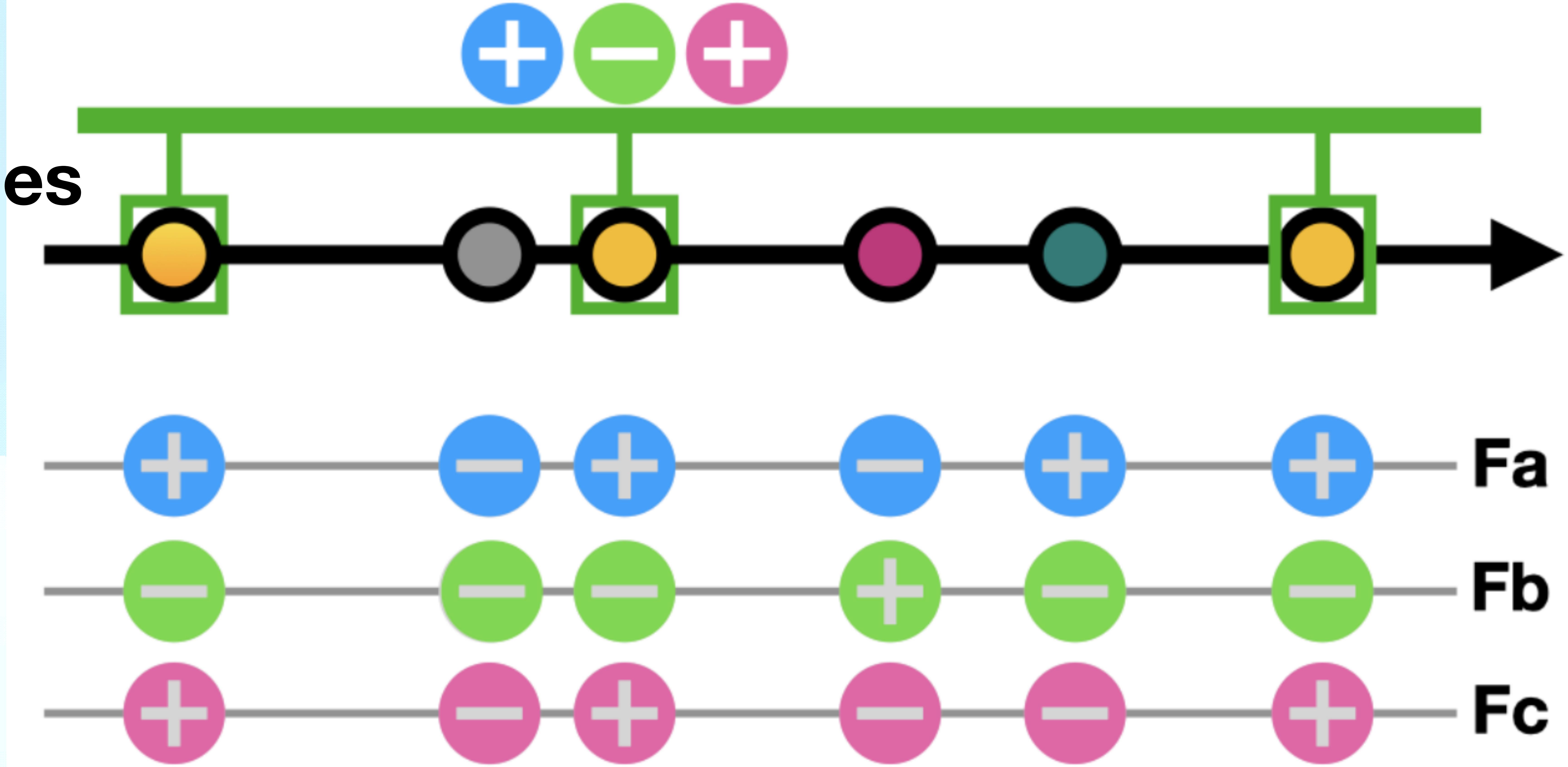
Homer Jay Simpson (born May 12 1956) is the main protagonist and one of the five main characters of The Simpsons series (or show). **He** is the spouse of Marge Simpson and father of Bart, Lisa and Maggie Simpson. **Homer** is overweight (said to be ~240 pounds), lazy, and often ignorant to the world around **him**.

Windows onto Variation

Inference from Corpus

- Use the corpus as its own model
- Identify which categories of contexts are likely to include all tokens chosen under the same influences
- true variation = the distribution of possible forms *for the same context*
- Ellison & Same (2022)

Categories



The probability of using a particular REF is its relative frequency in the category.

Windows onto Variation

LLMs

Give a large language model the same experimental task.

What referring expressions does it add?

INSTRUCTIONS GIVEN TO CHATGPT:

In the following, you will be presented with a news text from an American newspaper. The text will talk about one specific subject, which can be, for example, a person, a company, a group, a country or a commercial product. In the text, all references to one subject are replaced by [REF] (e.g., REF1, REF2, REF3, etc). Your goal is to fill those gaps, referring to the main subject of the text, so that the text becomes easy to understand.

Always talk about the whole subject: so if the topic is *Mr and Mrs Smith* and you know they are a couple from London, then you can fill a box with "Mr and Mrs Smith", "the London couple", "this couple", "they", "Mr and Mrs Smith's", "the couple's" and so on. But a box is never about "Mr Smith" or "Mrs Smith" on their own.

Although we show you words like "he", "she", "it", "they" or similar, you do not have to use them. In the box you can put any way of identifying the subject. If the subject is "Joe Biden, the president of the United States" you can put "Joe Biden" or "Mr Biden" or "Joe Biden's" or "the president of the United States" or "the president" or "the president's" or "him" or "his", or any other way of identifying this person.

At the start of each text, you will see a line which gives you the subject of the text (e.g., subject: Margaret Thatcher (she/her)), and the thing or person you need to refer to as you fill out the boxes. To help you understand how to talk about the thing or person you will also have a helper sentence providing more information (e.g., helper sentence: Margaret Thatcher was the prime minister of the United Kingdom.).

Here is the text:

TEXT1:

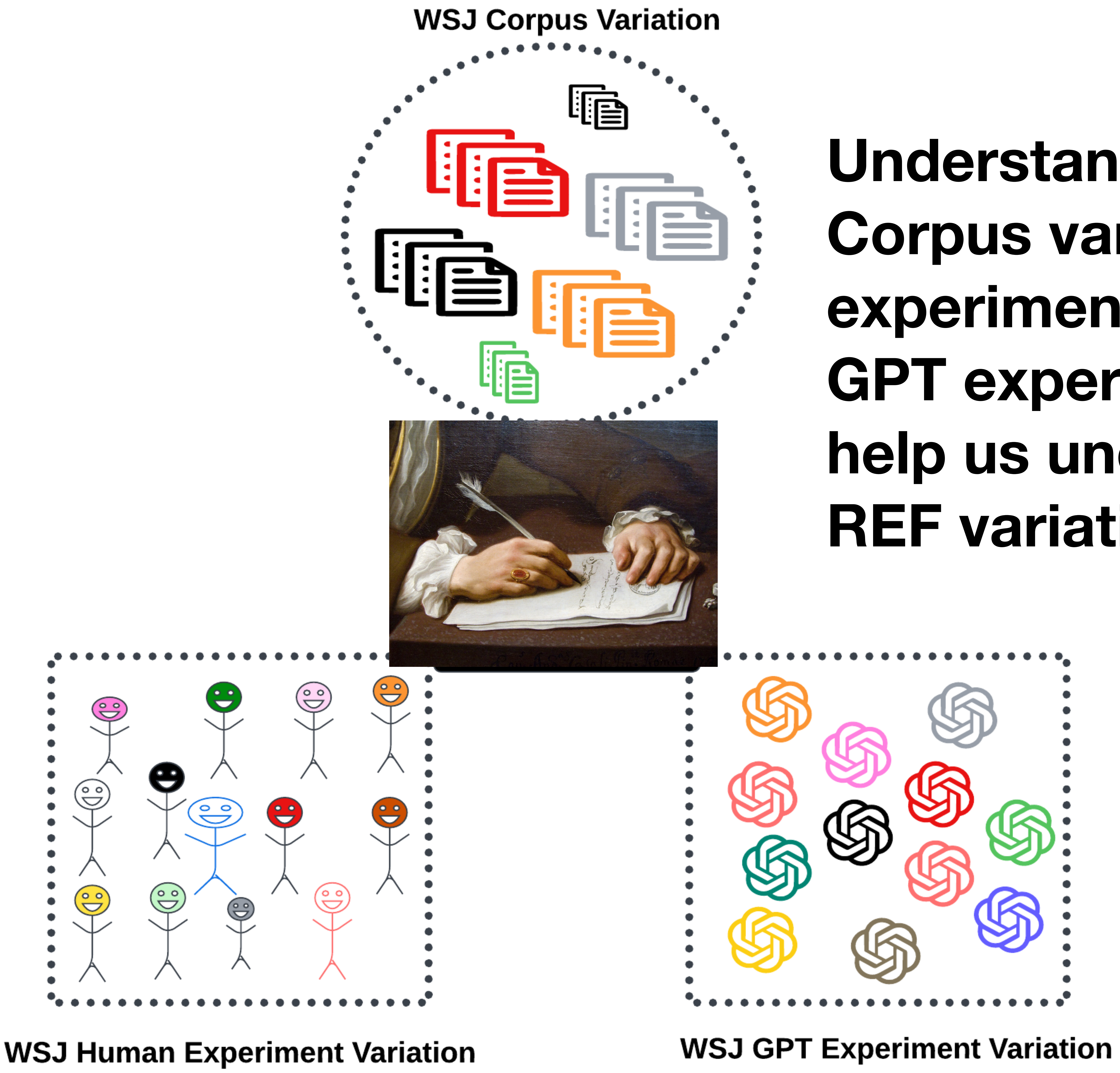
subject: Kenneth Roman (he/him/his)

helper sentence: Kenneth Roman is the 59-year-old former chairman and chief executive officer of the Ogilvy Group.

Just five months after Ogilvy Group was swallowed up in an unsolicited takeover , [REF1] said [REF2] is leaving to take a top post at American Express Co .

[REF3] abruptly announced [REF4] will leave the venerable ad agency , whose largest client is American Express , to become American Express 's executive vice president for corporate affairs and communications . [REF5] will succeed Harry L. Freeman , 57 , who has said he will retire in December . Mr. Freeman said in August that he would retire by the end of this year to take " executive responsibility " for an embarrassing effort to discredit banker Edmond Safra . American Express representatives apparently influenced the publication of

This Talk



WSJ Corpus Variation

WSJ Human Experiment Variation

WSJ GPT Experiment Variation

Understanding how Corpus variation, human experiments, and GPT experiments help us understand REF variation

Methods

for comparing distributions

$$JSD(d_1, d_2) = \frac{KL(d_1||d_{12}) + KL(d_2||d_{12})}{2}$$

$$d_{12} = \frac{d_1 + d_2}{2}$$

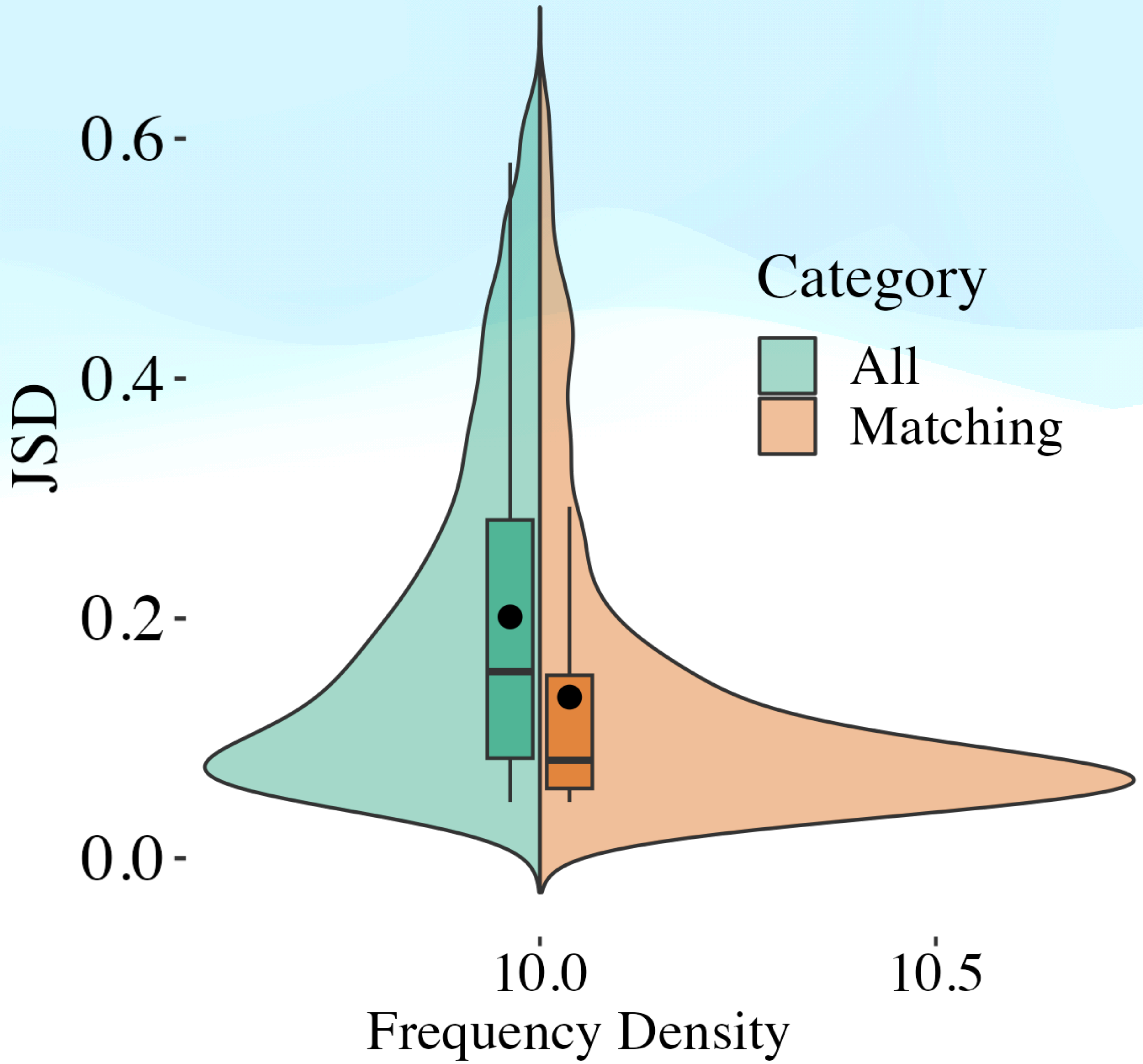
$$KL(d_2||d_1) = \sum_{f \in F} d^f \log \frac{d_2^f}{d_1^f}$$

- Jensen-Shannon Divergence (JSD) based on Kullback-Liebler divergence (KL) - shows difference between distributions

H1 - Our Categories are Predictive

Variation patterns are more similar within categories than between them

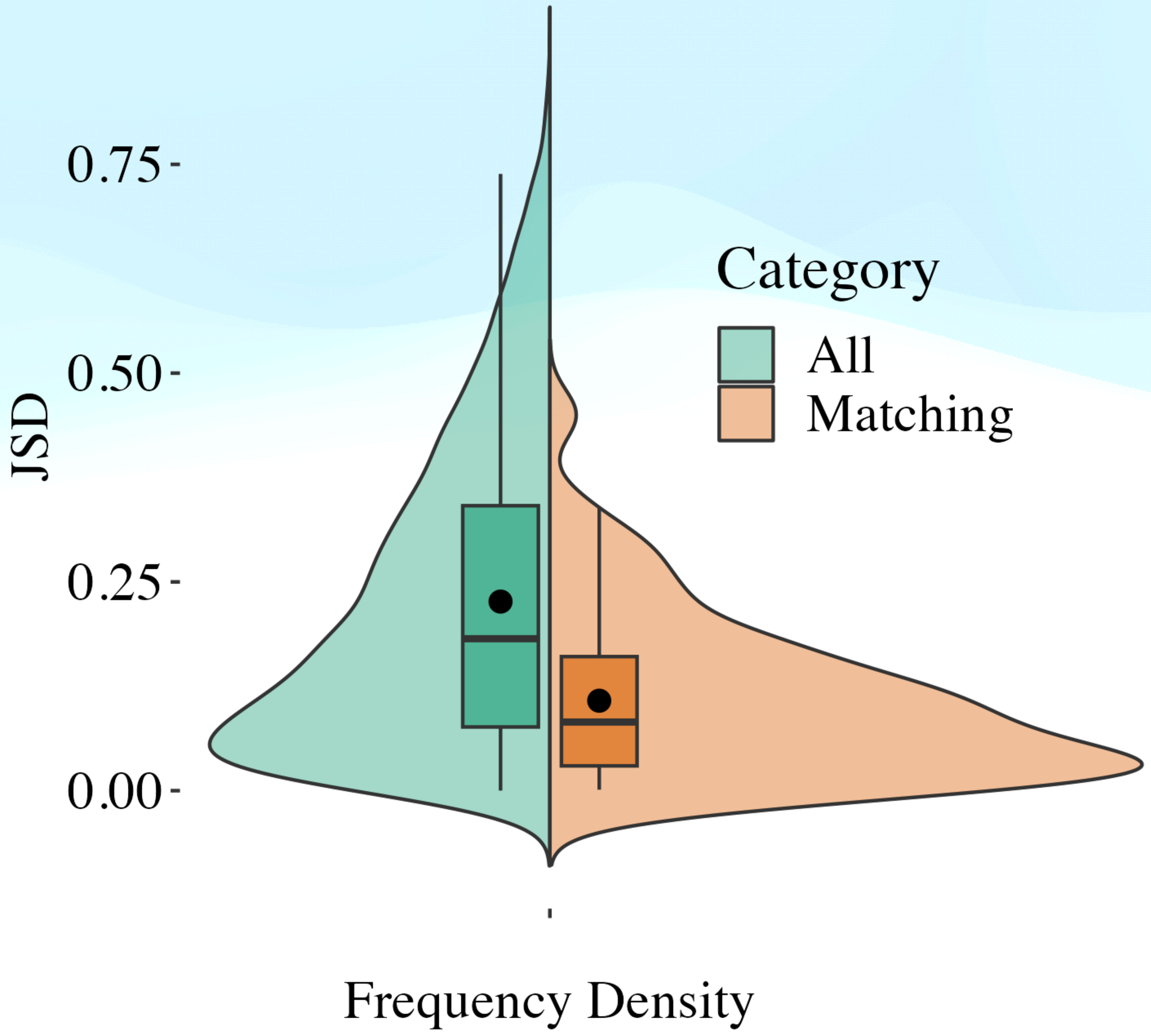
- experimental variation for each slot each story
- matching pairs randomly chosen from *matching* categories vs random pairs over *all* options
- matching pairs show much smaller difference between distributions of REFs for those slots



H2 - Corpus & Experimental Aligned

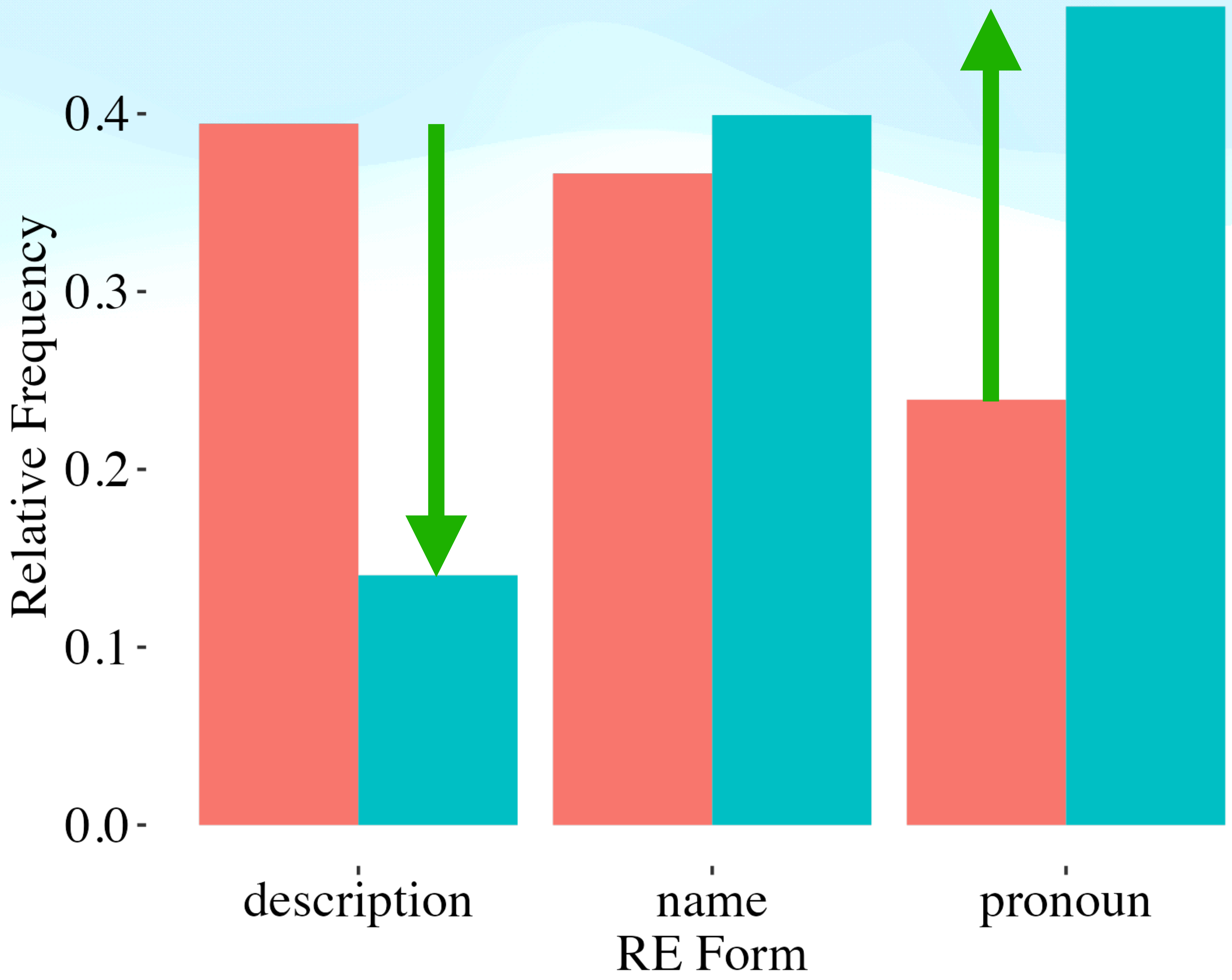
In-corpus and experimental variation are more similar when aligned

- JSD between category matched experimental and corpus distributions of variation
- either **with matching category**
- or **randomly matched**

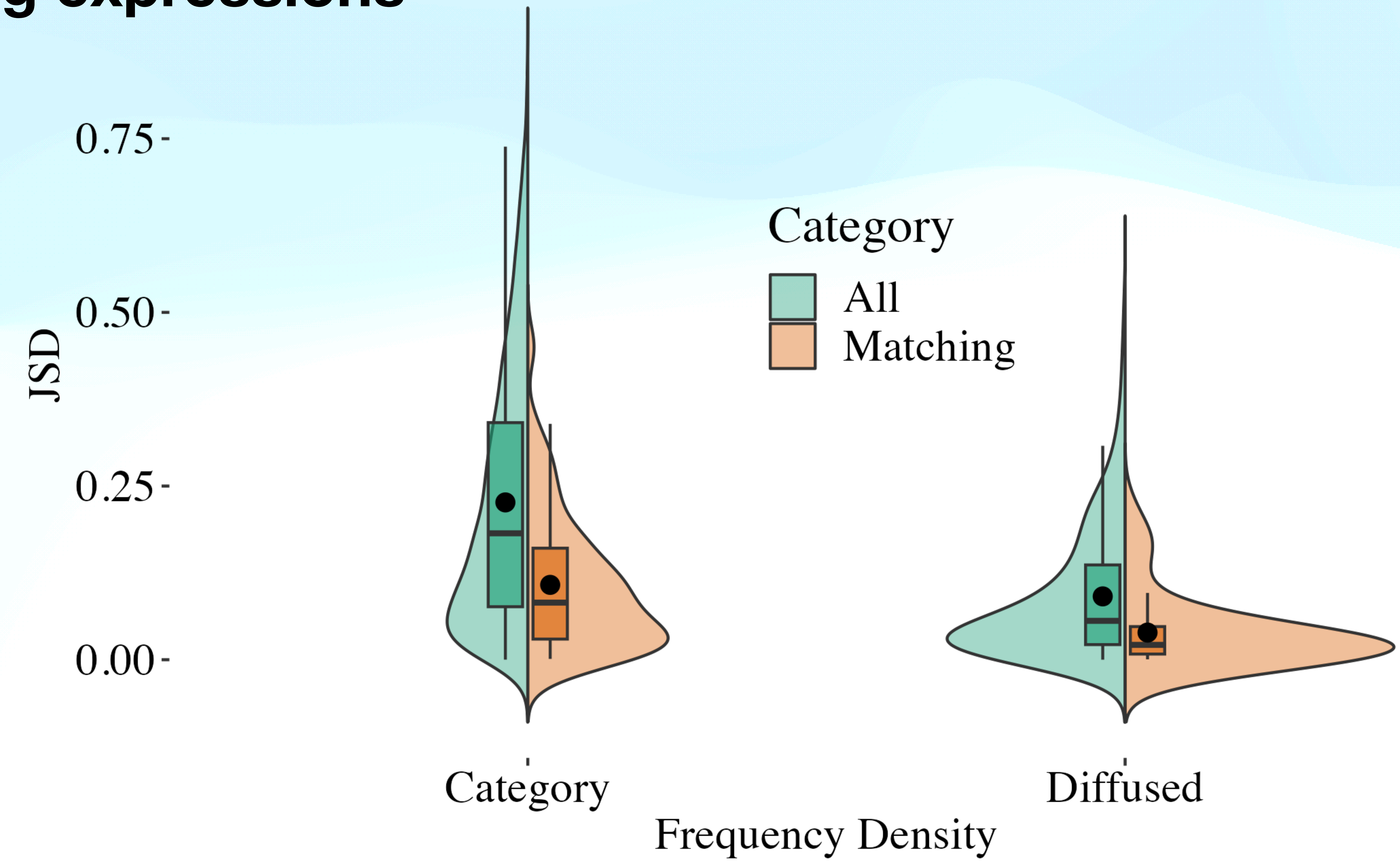


H3 and H4

In experiments, participants produce:
more pronouns
fewer descriptions

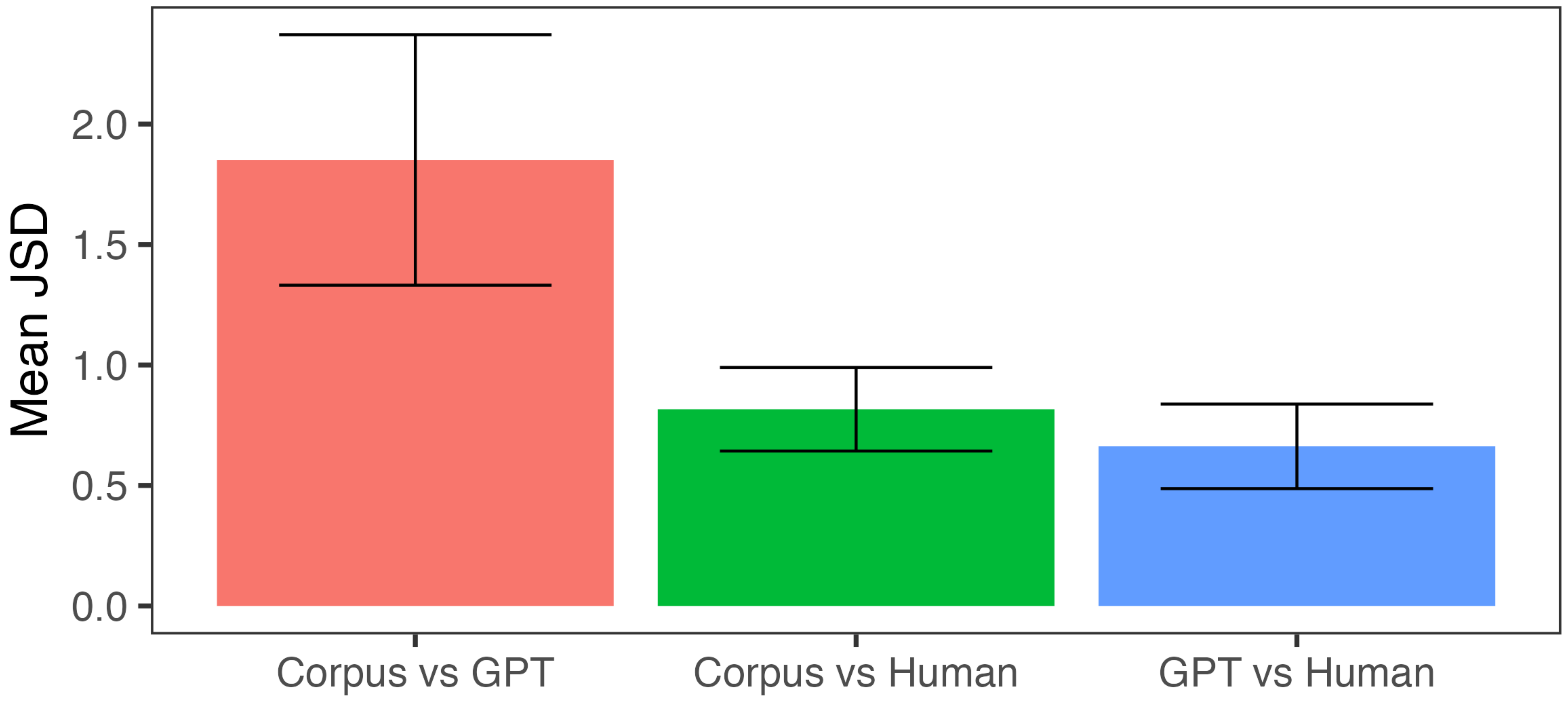


H5 Human experimental participants prroduce noiser distributions of referring expressions



H6 GPT REF distributions would be more similar to those from the Corpus than those from Experiments

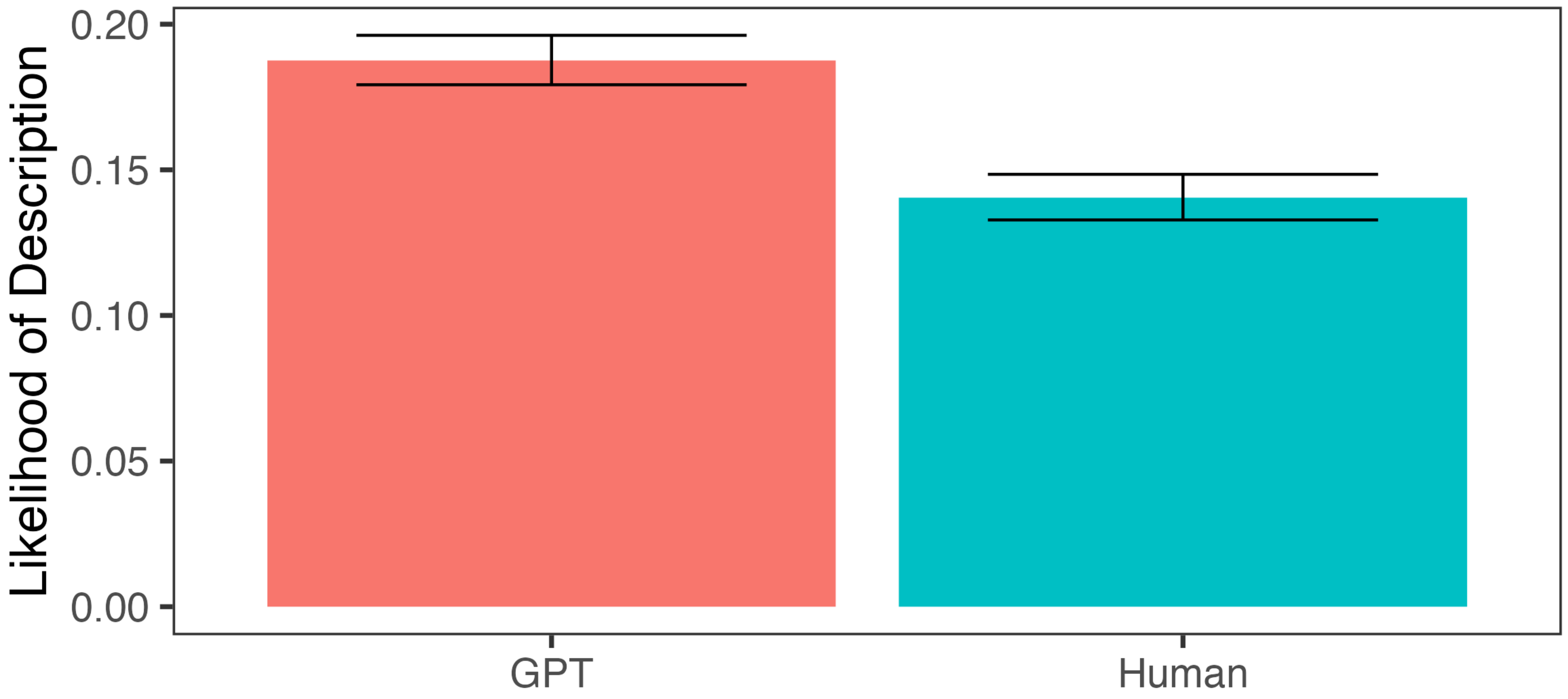
- We expected this because both are inferences being drawn from lots of textual data
- Instead, GPT variation distributions are more similar to humans in experiments, than to the corpus-based distributions



H7 GPT will produce more descriptive REs than experimental participants do

We expect humans to react to the text after understanding it, without exerting much effort in audience design.

The LLM is more likely to write for an audience (reflecting its training data), and so more likely to have descriptive referring expressions.



Conclusion

- real-time choice in REF is hard to capture
- the categories inferred from the corpus are reflected in experimental response
- experimental studies are better at predicting corpus distributions than LLM models

Thank you for your Attention