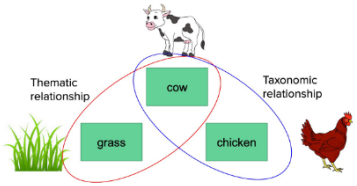
Re-Examining Cross-Cultural Similarity Judgments Using Lexical Co-Occurrence

Khuyen Le, Alexandra Carstensen, Michael C. Frank

BACKGROUND:

- **Cross-cultural differences in reasoning about similarity** have been attributed to **different notions of similarity** (Ji et al, 2004).



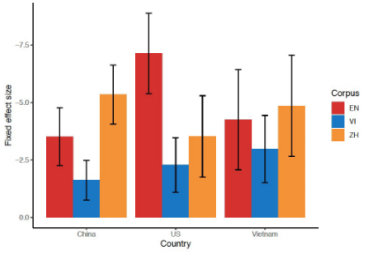
- Alternatively, these differences could be shaped by cross-cultural variation in the **statistics of the environment.**
- While statistics of everyday experience are difficult to measure, word co-occurrence can provide a rough proxy.

METHODS

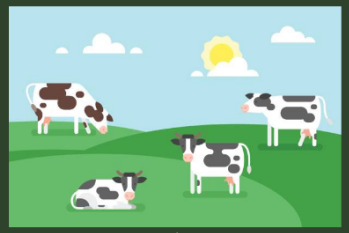
1. Participants from the US, China, and Vietnam (n=200, 200, 199 respectively) made 2AFC similarity judgements for 105 triads (as above).
2. Used linear mixed-effects models to test whether language-specific corpus co-occurrence (English, Mandarin, Vietnamese) can predict similarity judgements.

RESULTS

- We replicate differences in similarity judgments between the US and China, but these do not extend to our US-Vietnam comparison.
- Language-specific co-occurrence best predicted similarity judgments in China and the US.



Cross-cultural differences in **reasoning about similarity** may reflect **variation** in the **statistics of the environment** across cultural contexts.

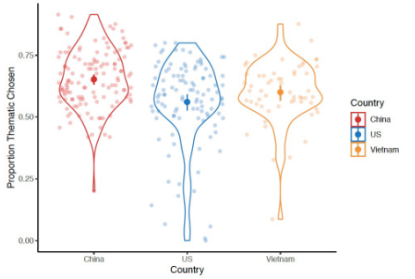


For instance, "cow" and "chicken" co-occur more frequently than "cow" and "grass" in a corpus of English, and participants in the US preferentially match cow and chicken in our similarity task, while the converse is true for Vietnamese and Mandarin Chinese corpora, as well as matching by participants in Vietnam and China.

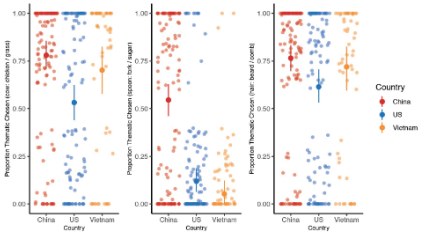
Take a picture to download my thesis

VARIATION ACROSS CULTURES

Previous research found that Chinese participants preferred thematic matches, while European Americans preferred taxonomic. We **replicate** this effect of country in our US-China comparison but **do not extend** it to the US-Vietnam comparison.

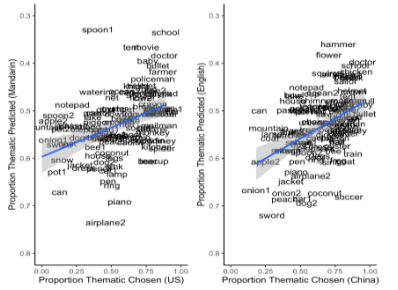


This effect varies across stimuli and might depend on the nature of the comparison.



UNIVERSAL TENDENCIES

We find strong evidence for **consistency** across the three groups: sizeable correlations across the corpus models, highly similar responding in the experiment, and correspondingly **high fit in cross-language comparisons** between models and data.



Language and Cognition at Stanford University