

What is more democratic, a stone or a feather?



UNIVERSITÄT
MANNHEIM



UNIVERSITY OF MANNHEIM

EXPERIMENTAL PSYCHOLOGY LAB

Predicting non-sensical choices using high-dimensional vector representations obtained from a semantic space model

64th TeaP 2022
March 20 - 23
Cologne, Germany



Introduction

Word2Vec

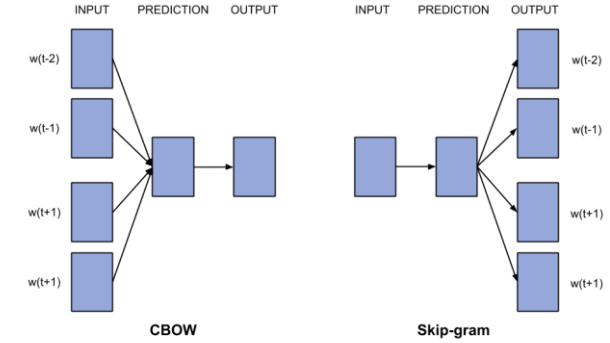
- is a vector-based semantic space model, where words are represented as high dimensional vectors (Mikolov et al., 2013)
- has been used to predict participants behavior in a variety of tasks:
 - Associative judgment (Bhatia, 2017)
 - Psycholinguistics (Mandera et al., 2017)
 - Similarity judgments (Pereira et al., 2016)
 - numerical estimation (Zou & Bhatia, 2021)

Research Question:

Can Word2Vec predict even non-sensical choices?



train



get

Dim	elephant	mouse	large
1	0.02	0.24	0.04
2	0.31	0.00	0.14
...	-0.27	-0.10	-0.12
300	0.02	0.10	0.05

compute

similarity(elephant, large) = .15
similarity(mouse, large) = .09

Method

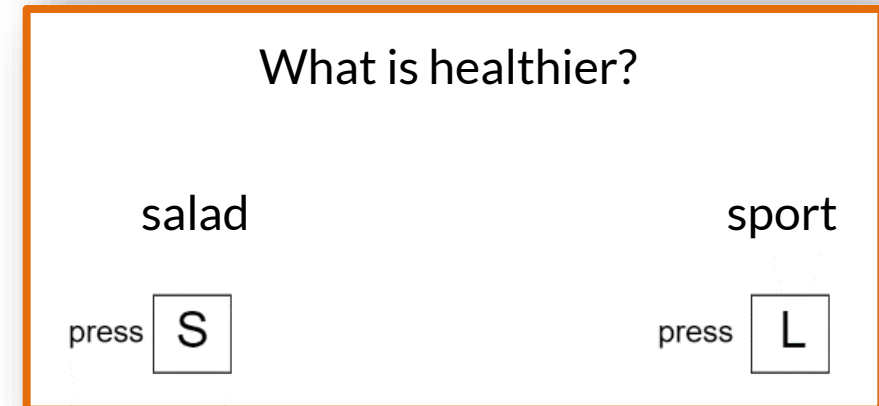
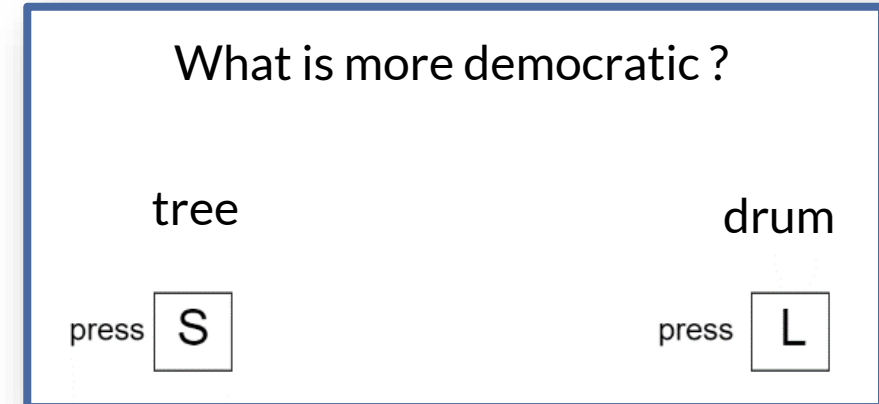
Sample:

- $N = 31$ participants, $M_{\text{age}} = 32.2$ ($SD = 10.6$), 74.2% female
- Participants were recruited via Prolific

Materials:

- 10 comparison dimensions [“democratic”, “expensive”, “healthy”, “undemocratic”, “sick”, “optimistic”, “rainy”, “hot”, “inexpensive”, “cold”]
- 6 word pairs per dimension
 - 4 word pairs, whose words could not be meaningfully rated on the dimension [*meaning = 0*]
 - 2 word pairs, whose words could be meaningfully rated on the dimension [*meaning = 1*]
- 2 attention check word pairs (clear correct choices)

→ 62 trials in total

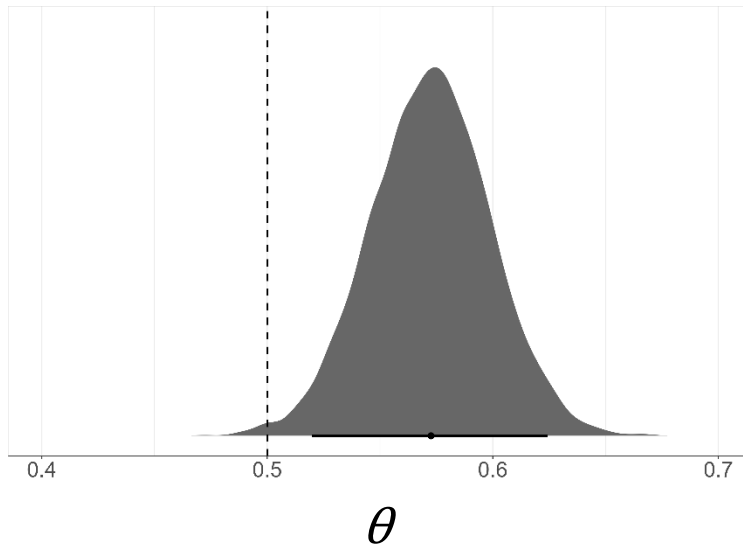


Measuring:

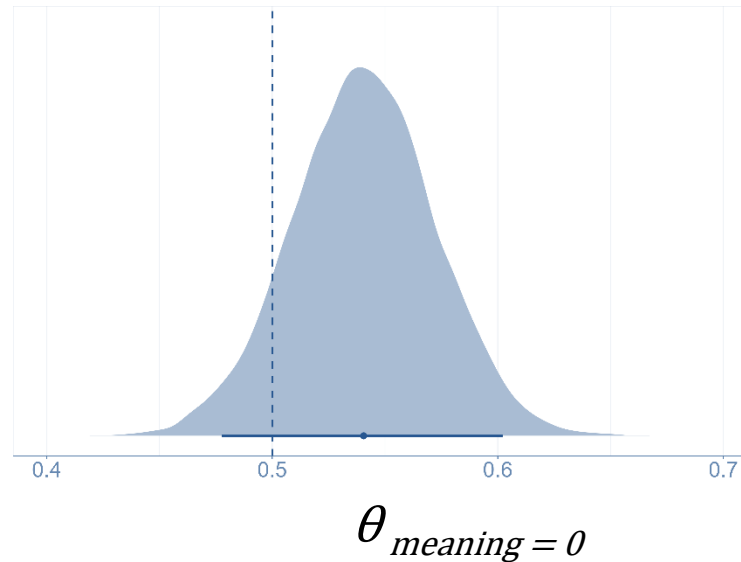
- Correspondence between the predictions of the vector-based semantic space model and the choice responses of the participants

Posterior Distributions of Model Coefficients

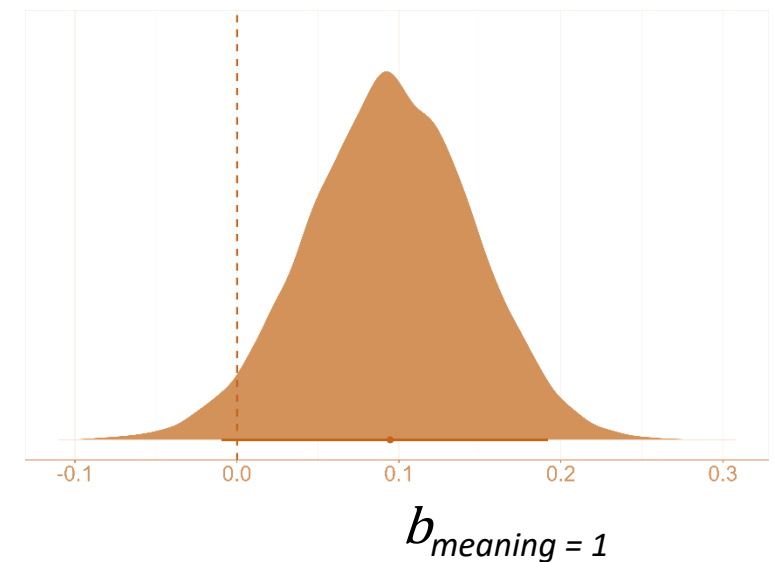
H_1 : The overall match rate between model predictions and participants responses θ is higher than chance level (i.e., $\theta > .5$).



H_2 : The match rate for the non-meaningful word pairs alone is still higher than chance level (i.e., $\theta_{meaning=0} > .5$).



H_3 : The model match rate is higher for meaningful than not meaningful word pairs ($b_{meaning=1} > 0$).



Data were analyzed using Bayesian hierarchical logistic regression models with random effects for participants and items

- Effects by

