Comparing the Multidimensional Mental Representations of Object Images and Object Nouns

Laura M. Stoinski[1,2,3], Tonghe Zhuang[4], Chris I. Baker[5] & Martin N. Hebart[3,4,6]

1 University of Leipzig, Germany; 2 International Max Planck Research School (IMPRS CoNI);
3 Max Planck Institute CBS, Leipzig, Germany; 4 Justus-Liebig-University, Giessen, Germany;
5 National Institute of Mental Health, Bethesda, MD 20814, USA; 6 Center for Mind, Brain and Behavior, Universities of Marburg, Giessen and Darmstadt

BACKGROUND

- We can think of semantics as composed of different dimensions
- Previous work showed that object images can be differentiated by a set of interpretable dimensions [1]
- These dimensions may vary for object words, which are not related to meaningful visual input

What are the core dimensions underlying object-word representations?

How do these dimensions differ from image-derived dimensions?

METHODS

- 1388 diverse, unambiguous object words, sampled from the **THINGS database** [2]

Word similarity task

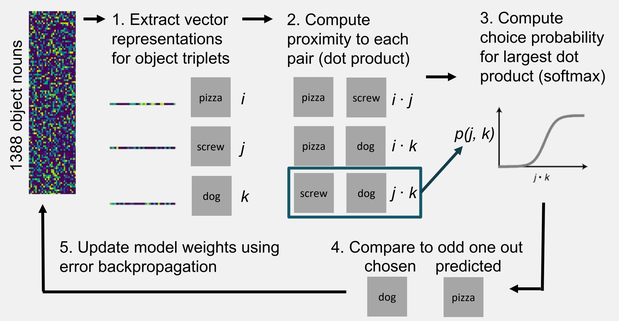
Large-scale online crowdsourcing, 5015 individuals (post exclusion)

Which is the odd-one-out?

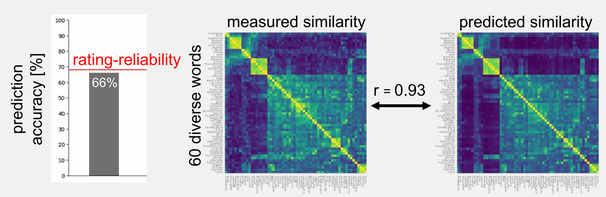
soup | jeep | snake

- ~1.3 mio. random combinations
- 1000 noise ceiling triplets (25-40 samples)
- All triplet combinations of 60 words

SPoSE model: computational model trained to capture similarity choices [1]

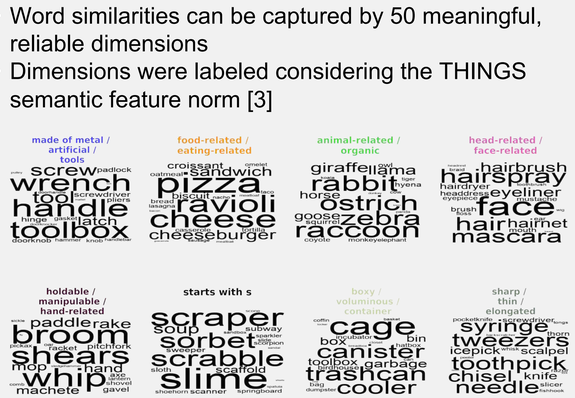


- Stable and predictive model of similarity choices

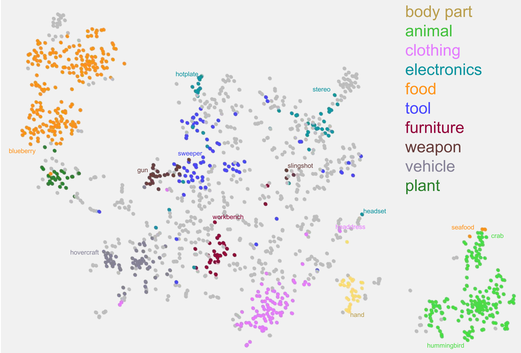


What are the core dimensions underlying object-word similarities?

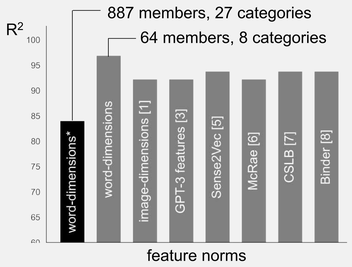
- Word similarities can be captured by 50 meaningful, reliable dimensions
- Dimensions were labeled considering the THINGS semantic feature norm [3]



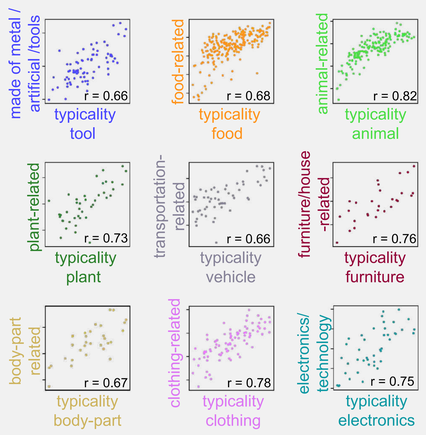
Do these dimensions capture known semantic relationships?



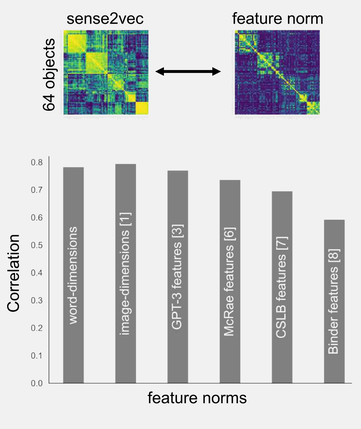
- Word-derived dimensions best predict memberships to **higher-level categories** [2, 4] (cross-validated nearest-centroid classifier)



- **Typicality** [4] of higher-level category members correlates with their loading on category-related dimensions → word-derived dimensions capture the richness of higher-level categories

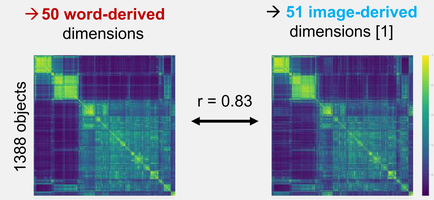


- **Sense2vec** [5] representations are best captured by image- and word-derived dimensions (RSA)



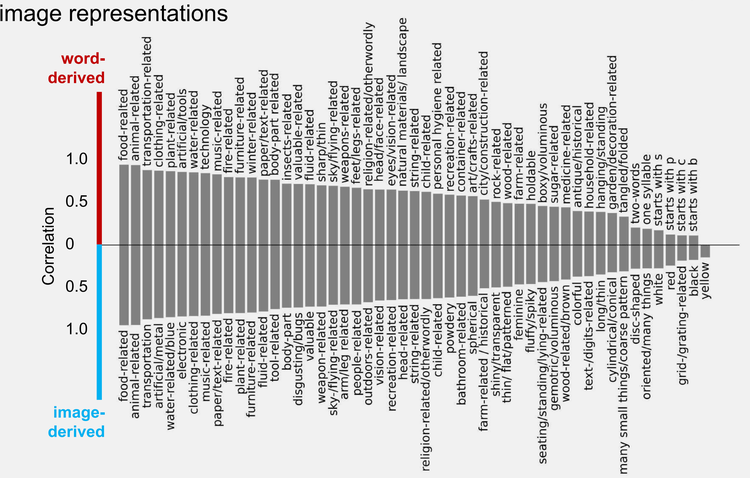
How do image and word representations differ?

- Word-representations capture much, but not all, of image-representations
- Differences remain after excluding dimensions tied to lexical-form and visual properties, or averaging image-derived dimensions across the estimated scores [9] for several images depicting the same concept



- without lexical-form dimensions: r = 0.84
- without dimensions related to visual properties: r = 0.84
- without both: r = 0.85
- when averaging image-dimensions across image examples: r = 0.84

- Correlations of word-derived dimensions with any image-derived dimensions and vice versa → color and texture dimensions are unique to image representations



CONCLUSION

- We identified **50 interpretable dimensions** underlying word similarity judgements
- These dimensions more effectively **capture higher-level categories** and known **semantic relationships** compared to other semantic norms
- **Word and image representations are very similar, but not the same** and their differences are not necessarily driven by **image-specific effects**
- Words are represented by only a **subset of visual-semantic** dimensions related to shape, but not color or texture

REFERENCES:
[1] Hebart, M. N., Zheng, C. Y., Pereira, F., & Baker, C. I. (2020). Revealing the multidimensional mental representations of natural objects underlying human similarity judgements. *Nature Human Behaviour*, *4*(11), 1173–1185. https://doi.org/10.1038/s41562-020-00951-3
[2] Hebart, M. N., Dickter, A. H., Kidder, A., Kwok, W. Y., Corriveau, A., Wicklin, C. V., & Baker, C. I. (2019). THINGS: A database of 1,854 object concepts and more than 26,000 naturalistic object images. *PLOS ONE*, *14*(10), e0223792. https://doi.org/10.1371/journal.pone.0223792
[3] Hansen, H., & Hebart, M. N. (2022). *Semantic features of object concepts generated with GPT-3* (arXiv:2202.03753). arXiv. http://arxiv.org/abs/2202.03753
M. T. Pilehvar and N. Collier, *De-Conflated Semantic Representations*. EMNLP 2016, Austin, TX.
[4] Stoinski, L. M., Perkuhn, J., & Hebart, M. N. (2023). THINGSplus: New norms and metadata for the THINGS database of 1854 object concepts and 26,107 natural object images. *Behavior Research Methods*. https://doi.org/10.3758/s13428-023-02110-8
[5] Trask, A., Michalak, P., & Liu, J. (2015). *sense2vec—A Fast and Accurate Method for Word Sense Disambiguation In Neural Word Embeddings* (arXiv:1511.06388). arXiv. http://arxiv.org/abs/1511.06388
[6] McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior research methods*, *37*(4), 547–559.
[7] Devereux, B. J., Tyler, L. K., Geertzen, J., & Randall, B. (2014). The centre for speech, language and the brain (cslb) concept property norms. *Behavior research methods*, *46*(4), 1119–1127.
[8] Binder, J. R., Conant, L. L., Humphries, C. J., Fernandino, L., Simons, S. B., Aguilar, M., & Desai, R. H. (2016). Toward a brain-based componential semantic representation. *Cognitive Neuropsychology*, *33*(3–4), 130–174. https://doi.org/10.1080/02643294.2016.1147426
[9] Kaniuth, P., Mahner, F. P., Perkuhn, J., & Hebart, M. N. (2024). *A high-throughput approach for the efficient prediction of perceived similarity of natural objects* (p. 2024.06.28.601184). bioRxiv. https://doi.org/10.1101/2024.06.28.601184

stoinski@cbs.mpg.de

PERCEPTION & ACTION: Vision

#F148