Similarity as a window on the dimensions of object representation

Bradley C. Love & Brett D. Roads
University College London
Experimental Psychology
26 Bedford Way
WC1H 0AP
London, UK

Corresponding author: b.love@ucl.ac.uk (B. Love)

Hebart et al. analysed 1.5 million human similarity judgments and found that natural objects are described by a small set of interpretable dimensions. Such large-scale analyses offer new opportunities to characterise how people represent their knowledge, but also challenges including scaling to even larger datasets and integrating accounts of semantic representation.

Keywords: similarity, knowledge representation, embedding spaces, multidimensional scaling, big data

Judging the similarity of two objects is relatively effortless to the point of appearing trivial. However, similarity judgments can reveal a great deal about how people represent and compare objects [1]. Although similarity has been studied for decades, only recently have cognitive scientists considered this question at scale analysing millions of ratings [1]. In some ways, this approach parallels work in machine learning where large datasets are used to develop and evaluate models. The ImageNet object recognition database, which consists of a large number of images with corresponding human judgments, is one such example [2].

The similarity at scale approach to elucidating how people represent and compare objects differs from prior approaches in which one group of people list features for objects and another group rate the features. These semantic norms can consist of thousands of features [3], which successfully capture human performance in a number of tasks such as semantic priming. However, one potential criticism of this earlier approach is that feature listings rely on self-report, which may bias toward features that are easier to verbalise.

One alternative to explicitly listing features is to infer representations in a more wholistic and indirect manner, such as through similarity judgments. Multidimensional scaling (MDS) [4] was a precursor to the large-scale similarity approach for inferring representation. MDS relies on people's similarity ratings but at a smaller scale. Here, quantity has a quality all its own. The dimensions one extracts from MDS or any other approach to finding low dimensional representations are determined in part by the stimuli considered. For example, if nothing edible is included in the stimulus sets, then no dimensions related to flavour will be extracted because such dimensions would not explain significant variance in the similarity ratings.

A recent study by Hebart et al. [1] addressed this issue by considering a representative and large stimulus set consisting of 1,854 objects. They used a triplet task in which participants chose which of three objects was the odd one out (i.e., least similar). With over a billion possible triplets, exhaustive sampling was not an option. Instead, the authors used an embedding approach in which gradient descent learning found a low-dimensional embedding (i.e., coordinates for the 1,854 items) that was as consistent as possible with the 1.5 million triplet judgments the authors collected. Much like word co-occurrence models in computational linguistics, information can effectively "spread" from one judgment to another (Figure 1a) to calculate an embedding without needing to sample all possible judgments.

The embedding the authors uncovered had a number of properties that speak to how humans represent objects. Interestingly, and in contrast to feature listing results [3], only 49 dimensions were required to represent the 1,854 objects and capture human performance on triplet judgments. Most of the recovered dimensions were interpretable. Examples of the 49 dimensions include metallic, food, tools, and sports. Moreover, many of these dimensions exhibited a typicality structure in which objects varied in meaningful ways along the dimension.

While valuable in its own right, this work will likely take on greater importance as others make use of the embedding space in their own research. The 1,854 images used were drawn from the THINGS database [5], which the spotlight's first and last author contributed, and for which they have collected fMRI and MEG data. Thus, the work of Hebart et al. adds an important component to an interrelated set of resources available to other researchers.

Although the work of Hebart et al. is impressive, no embedding is definitive because the particulars of a study, including the selected stimuli and judgment task, will shape the results. Additionally, details of the embedding procedure (i.e., algorithm) are critical. For example, the authors' assumption that embedding dimensions are non-negative played a large role in shaping the results. For non-negative dimensions, each dimension is essentially an absent vs. present feature. For example, on the wheel dimension, images without wheels would have low values (greater than zero) whereas images of motorcycles and cars would have larger values. By forcing values to be non-negative, dimensions can only add on top of one another to build representations. Without the non-negative constraint, dimensions could interact, much like how interfering waves in a pond can amplify or cancel one another. A mathematical analogy would be Fourier decomposition in which the dimensions are the sine waves at different frequencies and their amplitudes are the embedding values. One practical implication of the non-negative constraint is that more dimensions are needed to form the embedding and found dimensions will be more part-like (Figures 1b), which may make dimensions more interpretable. Without this constraint, an embedding could reside along a nonlinear manifold (Figure 1c). One interesting question is whether the brain's representations are more like Figure 1b or 1c. Non-negative representations may have advantages in terms of readout (i.e., communication between brain regions) whereas less constrained solutions may offer computational efficiencies. The brain may find an appropriate trade-off.

The authors' design choices led to an excellent solution, but one that differs from previous efforts. As mentioned, feature listing solutions are higher-dimensional [3]. Those solutions also contain category-specific features, such "has a beak" that are largely absent in the authors' solution. Moreover, many listed features, even for everyday objects, are relational or extrinsic in nature rather than intrinsic to the object [6], such as features related to function. People view thematically related objects (e.g., predator and prey, a man and his tie, etc.) as similar [6,7]. Likewise, embeddings based on people's real-world choices reveal that most dimensions are goal-relevant [8]. One interesting question is how different embedding spaces relate [9].

One particularly impressive aspect of the spotlight paper is the scale of the endeavour. To push even farther, judgements will need to be sampled non-randomly to focus on the informative judgments that reduce uncertainty in the location of objects within the embedding. We adopted such an approach using active learning to create an embedding space more than an order of magnitude larger using fewer similarity judgments [10]. We believe such ideas, combined with the types of contributions exemplified Hebart et al., will both help elucidate how people represent concepts and provide valuable resources to support allied endeavours.

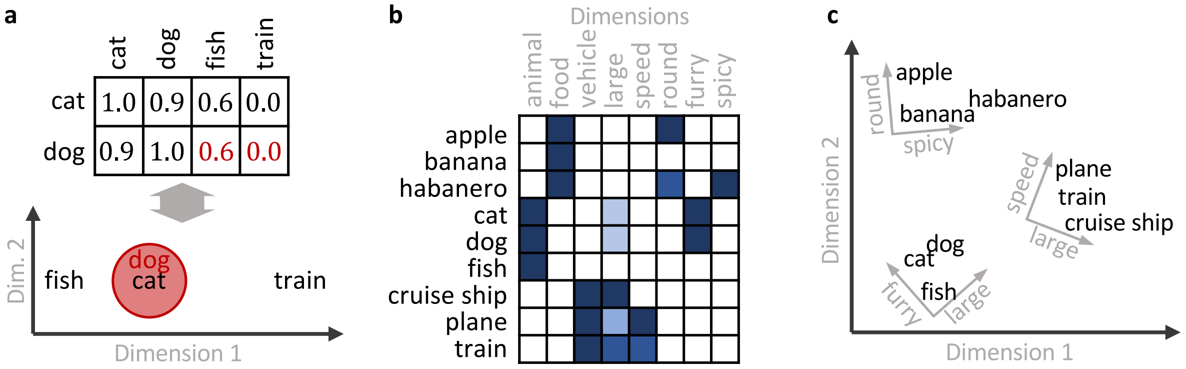

Figure 1: Inferring embedding spaces through similarity judgments. (a) Information contained in similarity judgments diffuses throughout the embedding. For example, consider part of a similarity matrix with unknown values shown in red -- the missing values can be guessed from neighbours. Given the constraints from neighbours, "dog" must reside within the red circle. (b) A visualization of a non-negative embedding where the stimuli (rows) and have sparse values across the different dimensions (columns). Darker cells indicate higher values. (c) An analogous non-linear embedding where "feature" dimensions are locally, but not globally, interpretable.

References

1. Hebart MN, Zheng CY, Pereira F, Baker CI: **Revealing the multidimensional mental representations of natural objects underlying human similarity judgements**. *Nat Hum Behav* 2020, **4**:1173–1185.

2. Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L: **ImageNet: A Large-Scale Hierarchical Image Database**. *CVPR09* 2009,

3. McRae K, Cree GS, Seidenberg MS, McNorgan C: **Semantic feature production norms for a large set of living and nonliving things**. *Behav Res Methods* 2005, **37**:547–59.

4. Shepard RN: **The analysis of proximities: Multidimensional scaling with an unkown distance function. Part 1**. *Psychometrika* 1962, **1**:125–140.

5. Hebart MN, Dickter AH, Kidder A, Kwok WY, Corriveau A, Van Wicklin C, Baker CI: **THINGS: A database of 1,854 object concepts and more than 26,000 naturalistic object images**. *PLOS ONE* 2019, **14**:e0223792.

6. Jones M, Love BC: **Beyond Common Features: The Role of Roles in Determining Similarity**. *Cognit Psychol* 2007, **55**:196–231.

7. Bassok M, Medin DL: **Birds of a feather flock together: Similarity judgments with semantically-rich stimuli**. *J Mem Lang* 1997, **36**:311–336.

8. Hornsby AN, Evans T, Riefer PS, Prior R, Love BC: **Conceptual Organization is Revealed by Consumer Activity Patterns**. *Comput Brain Behav* 2020, **3**:162–173.

9. Roads BD, Love BC: **Learning as the unsupervised alignment of conceptual systems**. *Nat Mach Intell* 2020, **2**:76–82.

10. Roads BD, Love BC: **Enriching ImageNet with Human Similarity Judgments and Psychological Embeddings**. *ArXiv201111015 Cs* 2020,