Subtle differences in the perceptual spaces of low-level features and objects

Suniyya A. Waraich[1] and Jonathan D. Victor[2]

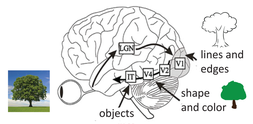
1. Weill Cornell Graduate School of Medical Sciences
2. Feil Family Brain & Mind Research Institute, Weill Cornell Medical College, NY

Weill Cornell Medicine
Graduate School of Medical Sciences

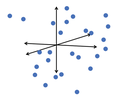
36.340
VSS 2022

Introduction

As signals pass through the network of visual areas, purely visual information is transformed into semantic information. But how does the representation of early-stage visual information differ from the representation of semantic information?



We study the mental representations of five distinct stimulus domains varying in their semantic content, using the framework of perceptual spaces.[1] A perceptual space is a mental representation in which points in a space denote stimuli and distances denote perceived dissimilarity.



Hypothesis: The mental representations of low-level features and semantic information have different geometric properties.

Methods

Using 5 stimulus domains varying in their level of semantic content (examples below), we ran parallel psychophysical experiments, with 9 subjects. We assessed the geometry of the representation of each domain by analyzing subjects' similarity judgments. All stimuli were based on a set of 37 familiar animals derived from WordNet.

Stimulus Domains

- Textures: fully scrambled textures of checks with colors taken from the image-like stimuli
- Texture-like Stimuli: texturized images of the animals[2]
- Image-like Stimuli: slightly pixelated images of the animals[2]
- Images: 37 unique recognizable images of the animals
- Words: the names of the animals

Data Collection:

- Subjects: (7F, 2M), VA: 20/20
- Image stimuli:
  - Image size: 2.25 deg
  - Check size (textures): ~ 13.3 arcmin
  - Diameter of the display: 12.2 deg

Data were collected via Zoom and remote screen control for 2 subjects and in-person for the remaining on a 13-inch laptop screen.

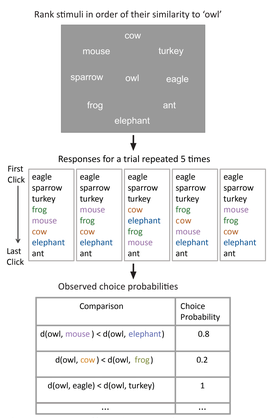


Figure A. *Collection and processing of similarity judgments*

Experimental Paradigm: In a typical experiment[3], a series of trials are presented in which 8 stimuli from one domain are shown around a central reference (Fig. A). The task is to click stimuli in the surround in order of similarity to the reference. There are 222 unique trials, in which each stimulus serves as the central item in 6 trials and is paired with each of the other 36 items at least once. These 222 unique trials are each repeated 5 times in the course of 10 sessions of 111 trials each. Subjects were debriefed and asked what strategies they used to gauge similarity.

Geometric Modeling: We derived Euclidean models (Fig. B) of perceptual spaces of 2 to 7 dimensions using a variant of multidimensional scaling, where the log-likelihood of observed choice probabilities was maximized by via gradient descent by adjusting the coordinates assigned to each stimulus.

Decision Model: Subjects' decision-making in each trial was interpreted as a set of independent, binary choices of the form "Is the distance between the reference and s_1 less than that between the reference and s_2?" for all pairs of stimuli in the surround. We modeled these decisions as the comparison of two distances with additive Gaussian noise representing errors in estimation.

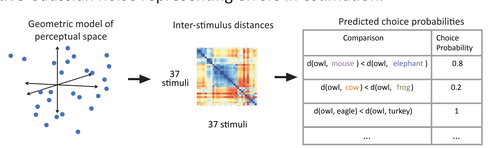


Figure B. *Distances between stimulus coordinates are used to account for model choice probabilities.*

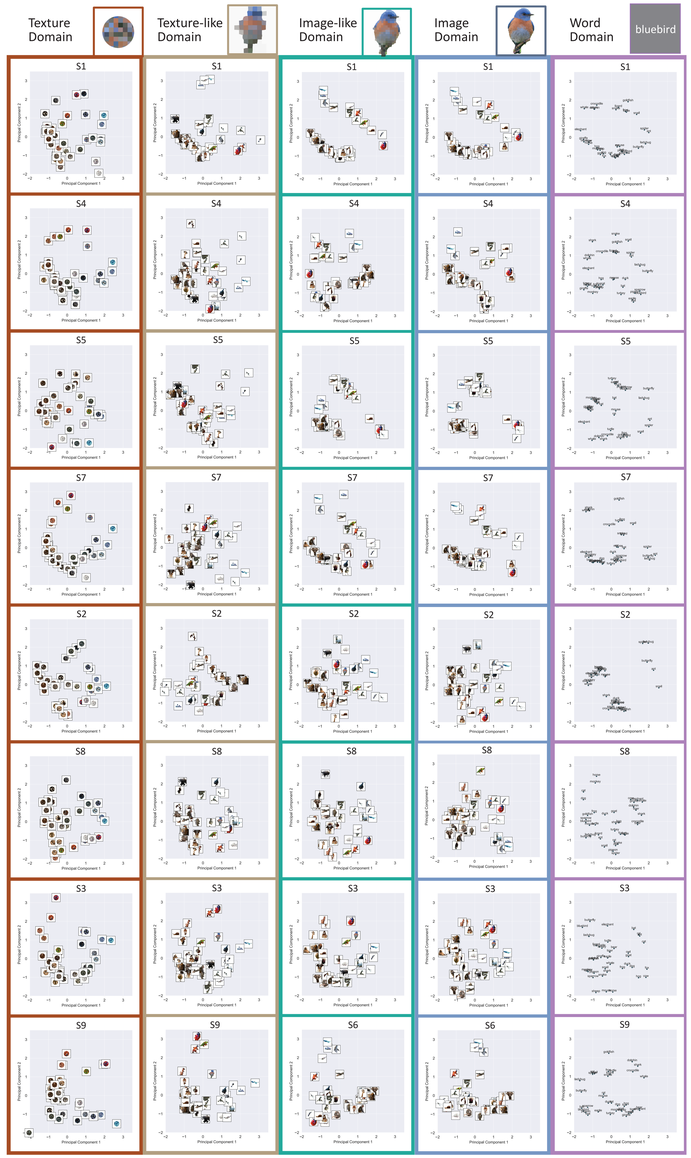
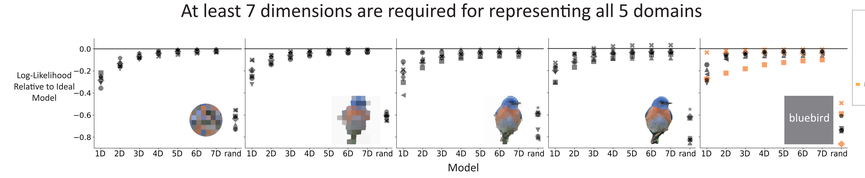


Figure C. *The projection of 5D coordinates onto the first two principal components, normalized by the variance along each component. Subjects S2, S8 and S3 used idiosyncratic strategies to perform the word experiment, judging similarity between animals by intelligence (S2), how much they would like the animal as a pet (S8) and phonetics (S3) respectively. The last row combines two subjects with partial data (see subject numbers in each panel).*

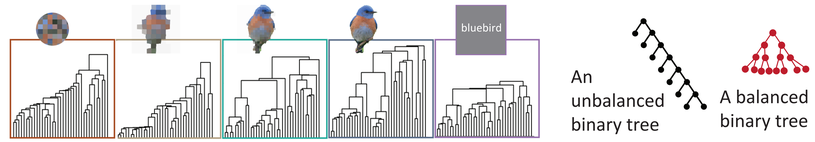
Results

At least 7 dimensions are required for representing all 5 domains

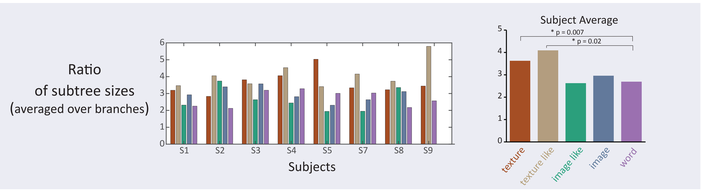


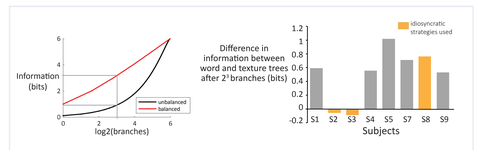
The log-likelihoods (LL) of observed choice probabilities by geometric models of varying dimensionality. LL's are plotted relative to the 'ideal' model LL, computed from the empirical probabilities.

Word space is represented by a more balanced tree



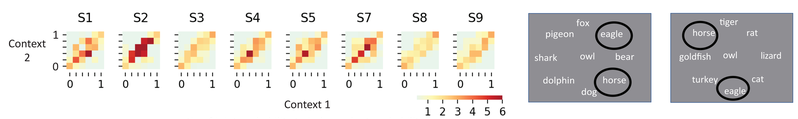
Hierarchical clustering was performed on the coordinates of the 5 dimensional model of each perceptual space for all subjects (dendrograms shown for Subject 1 on the left). Dendrograms of the word domains were the closest to being balanced, i.e., the sizes of the subtrees at each point of bifurcation were approximately equal (purple bars in the bar chart below).





More information is gained as one goes down a perfectly balanced tree than a completely unbalanced tree like the example above. This difference is greatest at 2^3 branches. For 6 out of 8 subjects, the word dendrogram contains more information than the texture dendrogram at the same depth from the root.

Similarity judgments were not dependent on context



The paradigm included trials in which the central reference appeared with a pair of stimuli in two different groupings. Judgments were highly consistent across contexts. Data shown are from the word experiment.

Conclusions

- All similarity spaces require at least 7 dimensions.
- Semantic space is more tree-like than the early-stage texture and intermediate texture spaces.

References

1. Zaidi, Q., et al. (2013). J Neurosci 33(45): 17597-602.
2. Achanta, R., et al. (2012). IEEE Trans Pattern Anal Mach Intell, 34(11): 2274-82.
3. Waraich, S.A., Victor, J.D., (2022). J. Vis. Exp. (181)

We would like to thank Mary Conte for her many helpful suggestions.
Support: NIH EY7977, Fred Plum Fellowship in Systems Neurology and Neuroscience
Contact: Suniyya Waraich at saw4003@med.cornell.edu