Emergent dimensions underlying human perception of the reachable world

VISION SCIENCES LABORATORY

Emilie L. Josephs, Martin N. Hebart & Talia Konkle

Introduction

Explicit similarity judgments can reveal the factors that shape how we reason about the world

What factors and distinctions characterize intuitive judgments of reachspace similarity?

Method

Approach: Capture the representational space of 990 reachspace images.

1. Simuli

- 990 Images: 3 images each from 330 categories
- Drawn from Reachspace Database (osf.io/bfyxk)
- Very wide sampling of reachspace views

2. Behavioral task

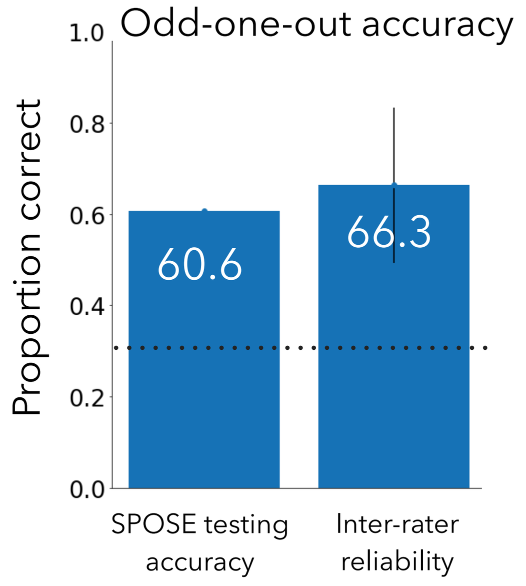
- Triplet similarity: "Which is the odd-one-out?"
- ~1.25 million trials on Mechanical Turk (0.8% of total possible triplets)
- 20 trials per HIT, no trial limit
- Stringent data quality checks enforced

3. Modeling

- Sparse Positive Similarity Embeddings (Hebart et al., 2021)
- Predictive model of similarity judgments
- Derives embedding for images: learns weights along inferred dimensions
- SPOSE model yielded 30-dimensional embedding

Model Validation

Does SPOSE embedding accurately predict similarity judgments?

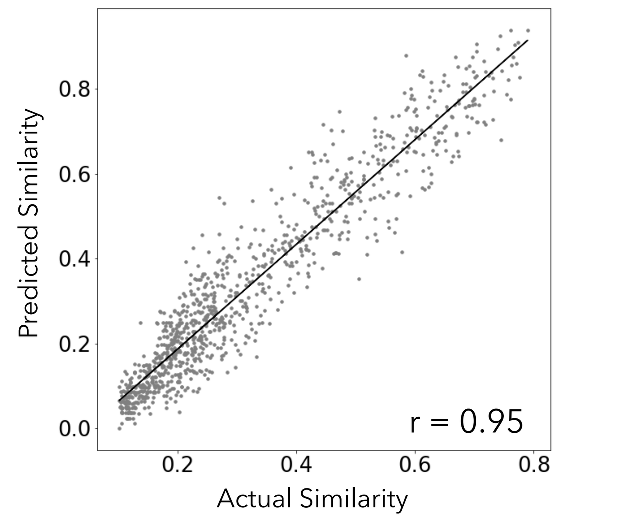


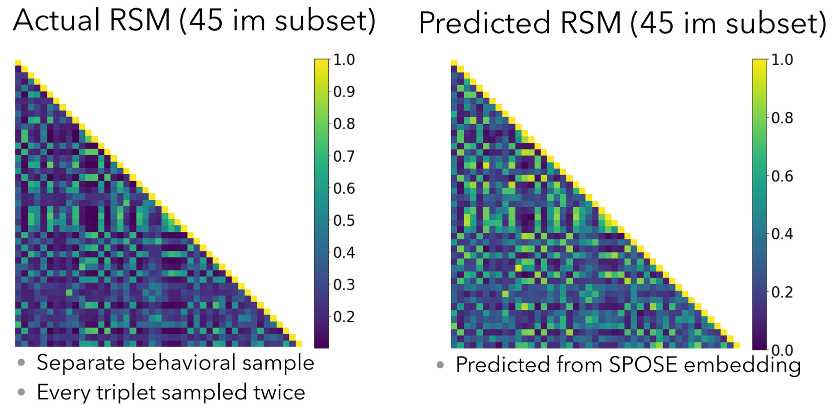
Inter-rater reliability: degree of agreement across 50 judgments, for 1000 triplets (separate behavioral sample)

Embedding captures 82.1% of explainable variation in individual trial outcome

Does the embedding capture the behavioral similarity space?

Corr between actual and predicted RSM in validation set of 45 images

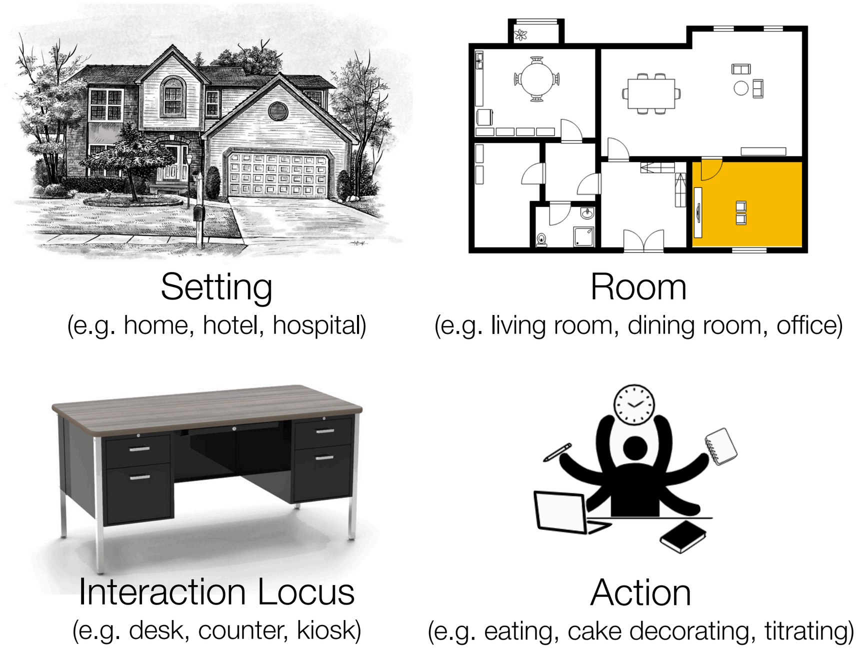


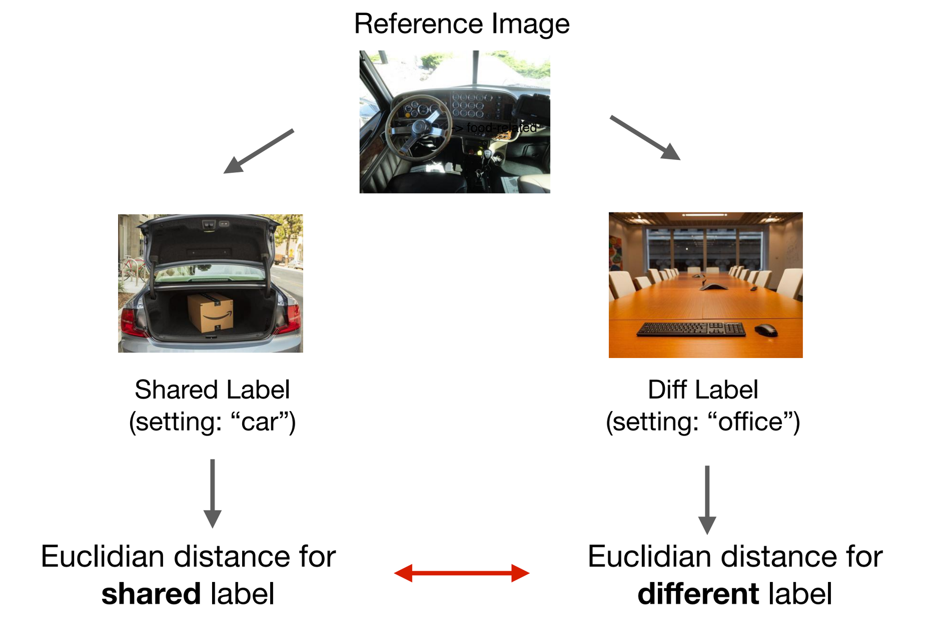


In a test set of 45 ims, embedding RSM is highly correlated with fully-sampled behavioral RSM.

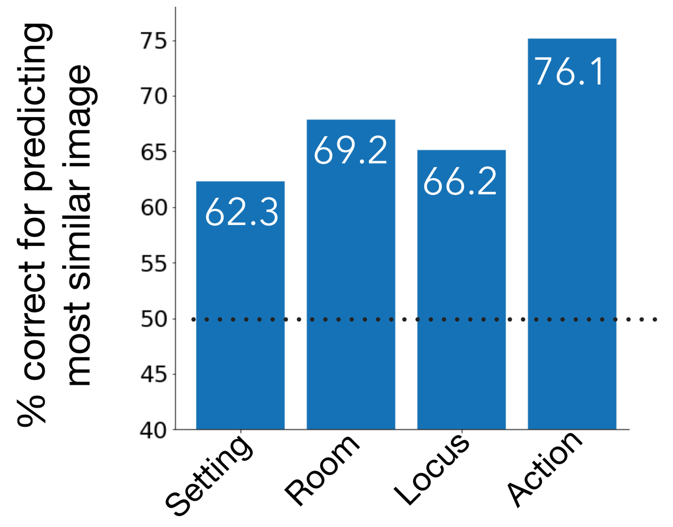
What factors underlie the similarity space?

Which factor is a better predictor of similarity: the location of a reachspace or the activity that it affords?



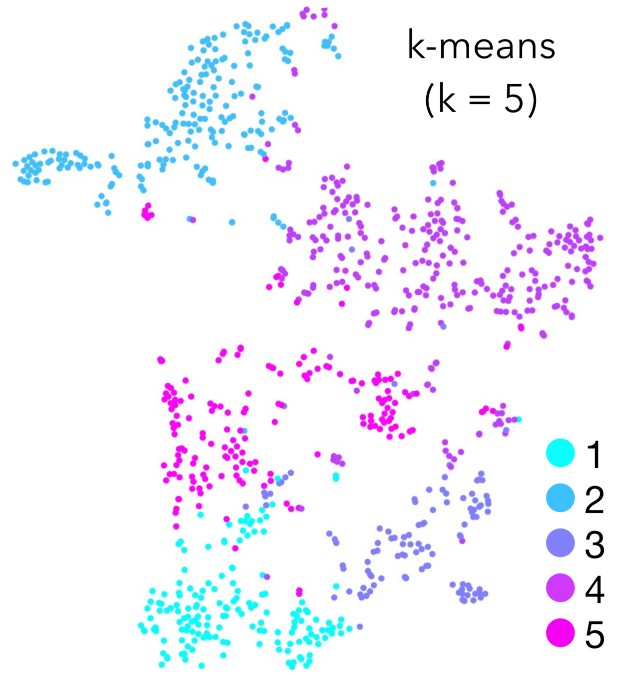


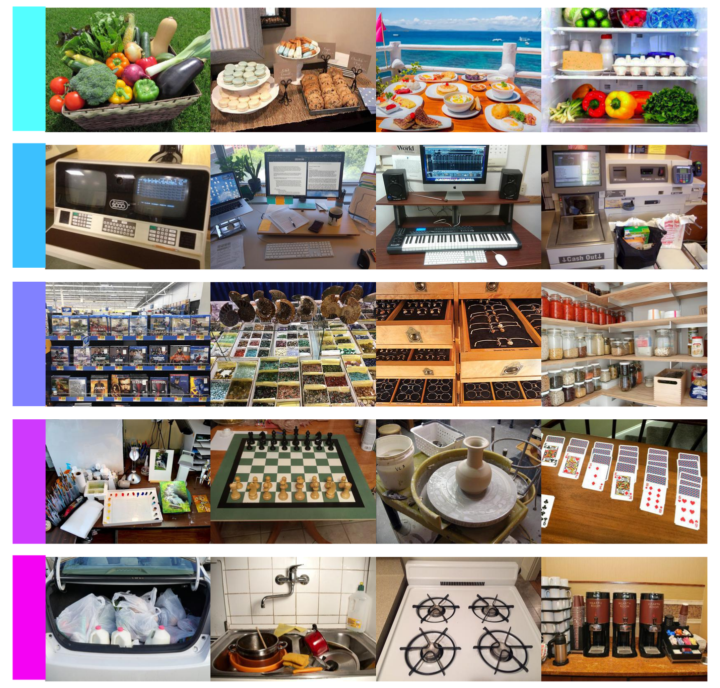
What percent of time were RSs which share a label judged more similar than RSs which didn't (Euclidean distance)?



Afforded action predict similarity judgments better than reachspace location.

What clusters characterize this space?





Assigning cluster labels:

- Experimenters generated possible labels for each cluster by inspecting images in them
- MTurk task: Indicate all images that fit a given label.
- Assess overlap between images that fit label and images in each cluster, with Adjusted Rand index.

	Food	Electronics	Retail	Storage/Display	Hobbies/Games	Chores	Liquids/Drinks
1	0.80	-0.10	0.06	-0.11	-0.10	0.11	0.18
2	-0.10	0.76	-0.09	-0.08	-0.09	-0.08	-0.09
3	-0.03	-0.09	0.67	0.44	-0.03	-0.01	0.01
4	-0.05	-0.04	-0.06	-0.01	0.36	-0.06	-0.05
5	-0.05	-0.08	-0.09	-0.01	-0.10	0.22	0.22

Do images with given label match any clusters?
Adjusted Rand Index (range: 0-1; 0=no correspondence)

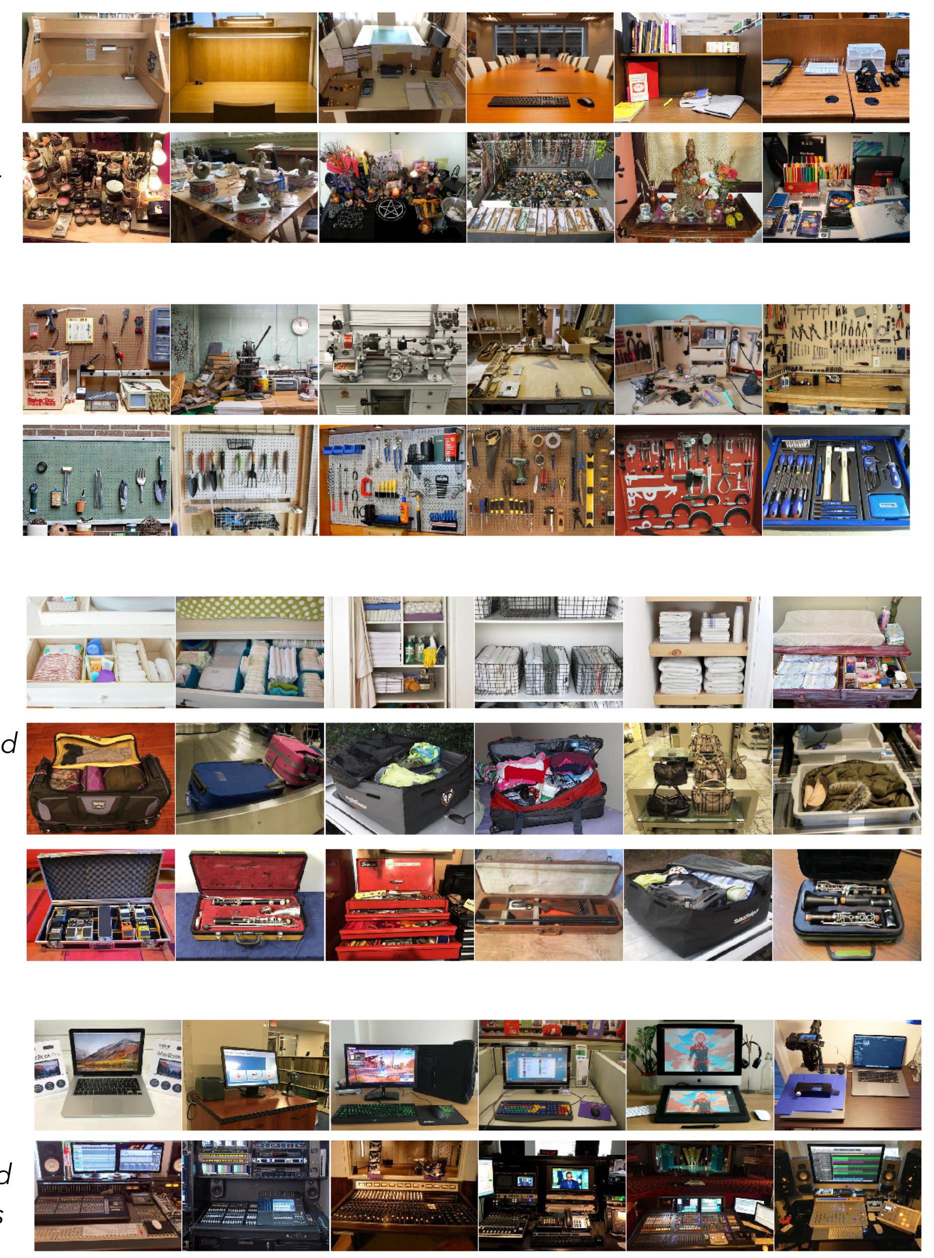
Clear clusters among these images for food-related, and electronics-related reachspaces.

Additional clusters for spaces related to display/storage, games/hobbies/crafts, and household chores/liquids.

Model Dimensions

A closer look at the embedding dimensions reveals nuanced distinction among reachspaces.
(Here, we spotlight 9 out of 30 dimensions)

low clutter
high clutter
workshop (active)
workshop (storage)
home storage
travel-related storage
portable storage
Consumer electronics
Specialized electronics



Conclusions

- How we reason about reachspaces relates to the things that we do in the space.
  - Echoes "design stance" toward objects and scenes (Keleman & Carey, 2007; Greene, Baldassano, Esteva, Beck & Fei Fei., 2016)
  - Suggests that the design stance shapes reasoning across a broad range of inputs and experiences.
- Can interpret clusters through action lens:
  - Food -> involves ingesting, hand-to-mouth kinematics
  - Electronics -> reasoning about hidden states, button or keypress kinematics
  - Non-active spaces -> occulomotor or simple grasping demands

Acknowledgments: This work was supported in part by NIH grant R21EY031867 to T.K., a Foundations of Human Behavior grant to E.J., and the Harvard Data Science Initiative Special Projects Fund.