Comparing facets of behavioral object representation: implicit perceptual similarity matches brains and models

Caterina Magri (cmagri@fas.harvard.edu)
Psychology Department, Harvard University, 33 Kirkland street
Cambridge, MA 02138 United States

Talia Konkle (tkonkle@fas.harvard.edu)
Psychology Department, Harvard University, 33 Kirkland street
Cambridge, MA 02138 United States

Abstract:

The similarity space of objects has been extensively used as a tool to relate representations among minds, brains, and models. However, the psychological construct of "similarity" is not well defined – objects can be similar in different ways. Here, we explored the similarity among inanimate objects, varying the instructions and task, and compared these to deep net representations and human brain responses. Specifically, we used a typical unguided sorting task in which participants drag and drop similar items nearby; a shape-guided sorting task, in which participants are explicitly instructed to arrange objects by shape similarity; and a pairwise-visual search task, in which participants have to find one target amongst others items, measuring similarity implicitly through reaction time. Our results show that (i) there are clear differences in the measured similarity space of objects across tasks, and (ii) the implicit similarity measured by visual search was better reflected in both deep net fits across all the early layers, and more extensively along the ventral visual stream. Broadly, these results highlight that different kinds of similarity can be manifest in different behavioral tasks, highlighting a rich space for elaborating the ways in which we explore representational matches between minds, brains, and models.

Keywords: shape; perception; object recognition

Introduction

We can effortlessly recognize thousands of objects, and we have knowledge about them—e.g. what they look like, how we use them, where we find them, and what they are for. Thus, there are many dimensions and properties along which objects can be similar to and different from one another. Measuring and relating object similarity spaces among behavioral measures, neural measures, and modeling responses to objects has been a powerful way to gain insight into the nature of visual object representations and its transformations along the ventral visual stream (e.g. Kriegeskorte et al., 2008, Yamins & DiCarlo, 2016).

While neural regions and model layers are typically assumed to have different representational similarity structures, that same assumption is not often applied to behavioral measures of similarity. A sometimes implicit assumption is that there is (only) one behavioral similarity space.

In the present work, we measured the similarity among a set of inanimate objects using three different behavioral methods, changing both the task and the instructions to emphasize varying degrees of perceptual similarity. Our goal was to understand how these behavioral similarity spaces relate to each other, as well as to both human brain responses and deep neural network (DNN) responses along their respective processing hierarchies.

Methods and Results

Behavioral Tasks

We measured object similarity spaces using three different approaches: 1) an unguided multi-arrangement sorting task (*sorting task* for short, Kriegeskorte and Mur, 2012), 2) a shape-guided sorting task that directs participants to mid-level visual feature information by instructing them to arrange objects by shape (given hypothesized importance of shape information in occipitotemporal cortex, e.g. Bracci & Op de Beeck, 2016; Long et al., 2018), and 3) a visual search task (Cohen et al., 2017).

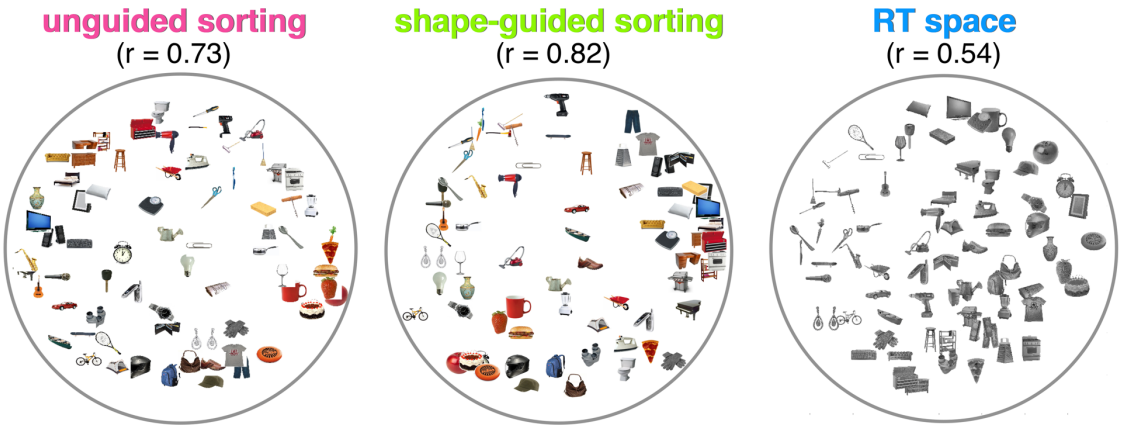

Figure 2: Exploring the behavioral spaces. MDS plots for the unguided sorting, the shape-guided sorting and the visual search tasks. R-values indicate split-half reliability estimates for each space.

For the sorting tasks, participants (n = 26 for unguided, n = 25 for shape-guided) were asked to arrange 72 images of inanimate objects in a circular arena (Kriegeskorte and Mur, 2012). This same task was run in two separate versions: an unguided version and a shape-guided version. In the unguided version we asked participants to arrange objects based on general similarity, without further specification of the kind of similarity to use. In the shape-guided version, participants were instructed to arrange objects specifically based on *shape*, and to avoid organizing them based on other properties (e.g. semantic category, color). The final output for both tasks is a dissimilarity matrix of distances for all the object comparisons for each experiment.

The visual search task (n = 1272) was an odd-one-out task, in which a participant had to find the unique object in a circular array with 5 copies of the distractor item (Figure 1). In this experiment, the unit of interest to measure object similarity is reaction time (RT): the faster the target is found among a particular distractor image, the more dissimilar the target is from that distractor. This experiment was conducted on using Mechanical Turk, and RTs were measured for all 2556 pairwise comparisons of 72 object images, Notably, in the first two tasks participants are asked to explicitly judge the similarity among different objects, while in the third task, the similarity measure is implicit in participants' reaction times.

Exploring the behavioral spaces

The reliability of all three tasks was moderately high, with a split-half correlation of 0.73 for the unguided sorting space, of 0.84 for the shape-guided sorting space, and of 0.55 for the search RT space.

We visualized the similarity space produced by the three tasks using multi-dimensional scaling (MDS; Figure 2). Qualitatively, visual inspection reveals that the unguided sorting space shows clustering by semantic categories (i.e. foods, musical instruments etc.) in contrast to the other two tasks.

Quantitatively, the shape-guided sorting space has comparable correlations with both the unguided sorting space (r=0.29) and the search RT space (r=0.34), which are both low, in the context of their internal reliability. The correlation between the unguided sorting space and the search RT space were much lower (r=0.12). Thus, these three similarity spaces seem to contain largely different information. In particular, even though the shape-guided sorting space and the RT spaces represent more closely shape-related features, they're not as correlated as one would expect and might reflect different aspects of the shape space.

Relating behavior and DNNs

How are these three facets of behavioral object similarity related to the representational transformations along the deep net layer hierarchy? To answer this question, for each layer of AlexNet (five convolutional and three fully connected layers), pre-

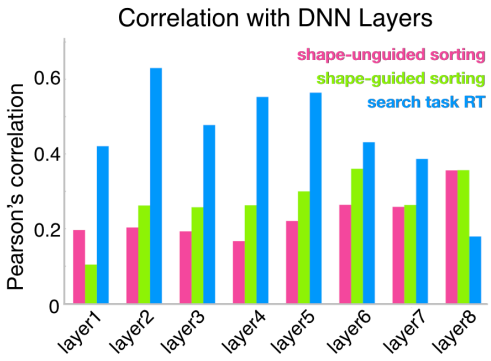

Figure 3: Relating behavior and DNNs. Weighted correlation between RDMs computed from AlexNet layers and behavioral RDMs.

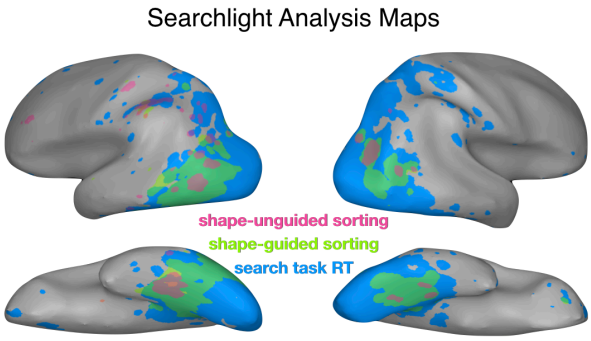

Figure 4: Relating behavior and brain: searchlight analysis. Surface brain map overlaying results of three separate searchlight analyses, one per behavioral task.

trained on ImageNet, we measured activations to each of the 72 items, and produced a representational dissimilarity matrix (RDM), using a Euclidean distance metric. We then computed the correlation between each layer's RDM and the three similarity spaces. Correlations were weighted by the individual tasks reliabilities so as to make them comparable across tasks. We find that the implicit task is most correlated with early to mid-layers of the DNN; only the last read-out layer 8 shows a stronger correlation for the two similarity spaces measured with the drag and drop task (Figure 3).

Relating behavior and brain

How do these three behavioral similarity spaces match neural similarity along the visual cortex? To explore this question, we scanned participants (n=11) as they viewed the same objects used in the behavioral studies, using a mini-block paradigm to measure reliable neural responses to each item (Konkle & Caramazza, 2014).

We then performed three separate searchlight analyses, correlating the behavioral similarity of each task with the neural similarity. These brain-behavior correlations were noise-corrected by both searchlight pattern reliability and task reliability. As a preliminary analysis step, we adopted a lenient threshold to investigate the broad similarity between neural and behavioral space. We restrict visualization to all voxels with brain-behavior correlations above r=0.25, overlaying them on the same surface map (Figure 4).

The searchlight results revealed that the visual search similarity space most extensively correlated with neural patterns throughout the visual system. Additionally, the shape-guided arrangement task was more strongly correlated with lateral and ventral OTC (lOTC, vOTC), and not parietal or early visual cortex. Finally, the unguided sorting task was the least extensive, correlating with neural similarity most strongly in anterior patches of lOTC and vOTC.

Conclusions

There are three main empirical findings of this paper. First, we found that there are different, reliable, behavioral similarity spaces among inanimate objects that can be elicited through different methods and instructions. Second, the similarity between pairs of items as measured implicitly by visual search speed has the strongest relationship with the similarity spaces in the layers of DNN models—all layers of an ImageNet-pretrained AlexNet (except the output layer) show the strongest relationship to search similarity. Third, an extensive swath of cortex is correlated with visual search similarity, with weaker and more restricted correlations evident in the arrangement-based similarity tasks.

One important distinction between these behavioral similarity spaces is whether similarity is judged implicitly or explicitly. Thus, one possible interpretation of these data is that the process of explicitly judging the similarity of two items may not give access to early perceptual representational similarity—that is, the kinds of representations that are evident in early and mid-stages of the visual system and of DNN representations. However, it is important to note that it is also possible that the arrangement-based method itself, rather than the explicit nature of the similarity judgments, is what prevented the more perceptual

similarity dimensions from manifesting in the measured similarity space.

Broadly, these results shed light on the nature of object representations in the visual system and in DNN models. First, these results show that there is an extensive swath of cortex that is correlated with visual search behavior, extending the work of Cohen et al, 2017 to item-level similarities only within inanimate object categories. Further, these results join an increasing body of literature suggesting that the visual system maintains a low-to-mid level perceptual representation, even at the more anterior stages of the hierarchy (e.g. Long et al, 2018; Baldassi et al., 2013). Finally, these results highlight that different kinds of similarity can manifest in different behavioral tasks, highlighting a rich space for elaborating the ways in which we explore matches between minds, brains, and models.

References

Baldassi, C., Alemi-Neissi, A., Pagan, M., DiCarlo, J. J., Zecchina, R., & Zoccolan, D. (2013). Shape similarity, better than semantic membership, accounts for the structure of visual object representations in a population of monkey inferotemporal neurons. *PLoS computational biology*, *9*(8), e1003167.

Bracci, S., Op de Beeck, H., 2016. Dissociations and Associations between Shape and Category Representations in the Two Visual Pathways. Journal of Neuroscience 36, 432–444. doi:10.1523/JNEUROSCI.2314-15.2016

Cohen MA, Alvarez GA, Nakayama K, Konkle T (2017) Visual search for object categories is predicted by the representational architecture of high-level visual cortex. J Neurophysiol 117:388–402.

Jozwik, K. M., Kriegeskorte, N., & Mur, M. (2016). Visual features as stepping stones toward semantics: Explaining object similarity in IT and perception with non-negative least squares. Neuropsychologia, 83, 201-226.

Konkle, T., & Caramazza, A. (2014). Object gist features capture the structure of neural responses to objects. *Journal of Vision*, *14*(10), 1292-1292.

Kriegeskorte, N., & Mur, M. (2012). Inverse MDS: Inferring dissimilarity structure from multiple item arrangements. Frontiers in psychology, 3, 245.

Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in systems neuroscience*, *2*, 4.

Long, B., Yu, C. P., & Konkle, T. (2018). Mid-level visual features underlie the high-level categorical organization of the ventral stream. *Proceedings of the National Academy of Sciences*, *115*(38), E9015-E9024.

Yamins, D. L., & DiCarlo, J. J. (2016). Using goal-driven deep learning models to understand sensory cortex. Nature neuroscience, 19(3), 356.