Determining visual shape features for novel object classes

Yaniv Morgenstern, Filipp Schmidt, Roland Fleming
Psychology Department, Justus-Liebig-Universität Gießen

The visual representation of shape reduces a high-dimensional input into a smaller set of more informative features. These features can span a range of abstractions from shallow features based on statistical summaries of images, to deep features related to the generative causes of the shapes. Here we examined the depth of the visual system's representation of shape by comparing human judgments of whether novel shapes appeared to belong to a common class with a range of models with different shape representations. Each shape class was based on a unique 2D base shape, formed by attaching parts of contours from different naturalistic shapes. We generated novel samples by transforming the base shape's skeletal representation (Figure 1A; Feldman and Singh, 2006) to produce new shapes with limbs that varied in length, width, spatial position, and orientation relative to the base shape. Multiple related classes were derived from each base shape using different distributions of parameter values. On each trial, observers judged whether the given *target shape* was in the same class as the given *context shapes* (either one or sixteen samples drawn from a particular shape class; Figure 1B). Target shapes were samples taken from the same shape class as the context or one of the 5 related classes (Figure 1C).

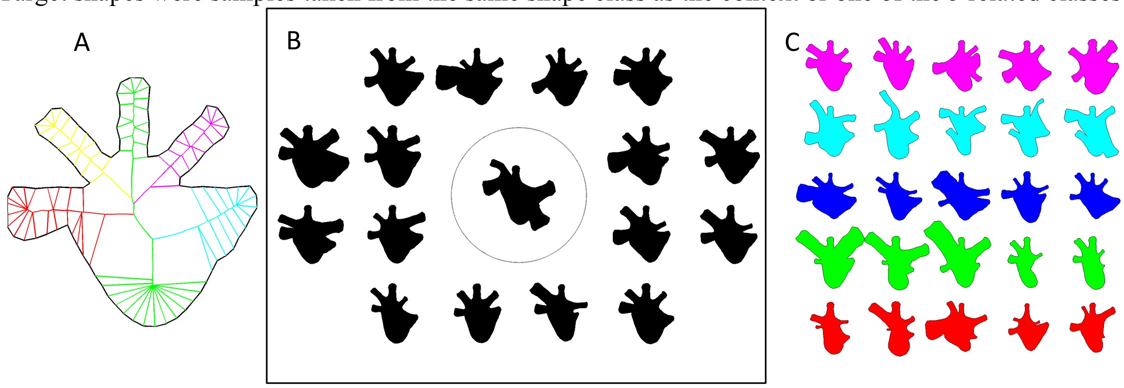

Figure 1. **Stimuli.** **(A)** The skeletal representation of a base shape was used to generate 6 different classes of shapes. (**B**) On each trial, observers were shown context shapes (1 or 16) from one class and asked to judge whether a target shape, presented within the circle, came from the same shape class as the context. (**C**) The target shape could come from the same class of shapes as the context or from other classes (colour coded).

Participants perform remarkably well given the ill-posed nature of the task. Models based on shallow features (Euclidean distance and shape area), and deep features (an ideal observer model with knowledge on the distribution of skeletal parts), were evaluated in terms of trial-by-trial consistency with the human data. In general, human responses indicated generalization beyond the context class and were best described by ideal and sub-optimal observer models suggesting that shape features for novel object classes are an abstract version of the underlying deep features (Figure 2).

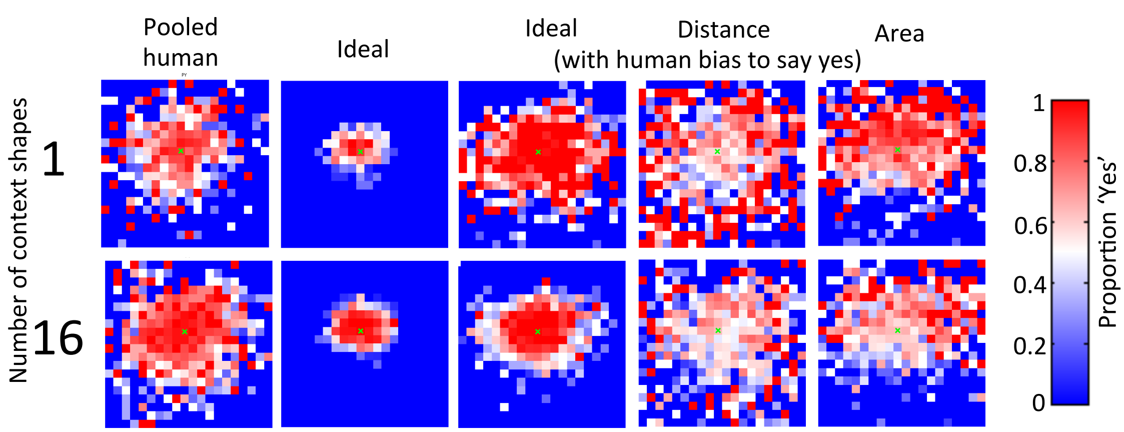

Figure 2. Human and model responses in MDS feature space. The bins in each subplot are arranged using MDS such that similar shape stimuli, in terms of the pairwise distances in the underlying parameter values used to generate the shapes, are closer together. The green 'X' (near the centre of each subplot) marks the location of the base shape. Each bin indicates the probability of a yes response (red = 100%, white = 50%, blue = 0%) for groups of shapes with similar values in MDS feature space to be perceived as part of the context shape class. The pattern of human results (pooled over 5 participants in each condition) is most similar to deeper models, like an ideal observer with human bias to say yes, than shallower models, like Euclidean distance or area.

References

Feldman, J., & Singh, M. (2006). Bayesian estimation of the shape skeleton. *Proceedings of the National Academy of Sciences*, *103*(47), 18014-18019.