THE PERCEPTUAL DIMENSIONS OF NATURAL DYNAMIC FLOW
Supplemental Content

Yaniv Morgenstern\textsuperscript{1,2*}, and Daniel J. Kersten\textsuperscript{1}

\textsuperscript{1} Department of Psychology, University of Minnesota,
75 East River Road, Minneapolis, MN, USA, 55455

\textsuperscript{2} Department of Psychology, Justus-Liebig-Universität Gießen,
Otto-Behaghel-Strasse 10F, Giessen, Germany, 35394

* Corresponding author: yaniv.morgenstern@psychol.uni-giessen.de
Figure S1. Multi-arrangement method of Kreigeskorte and Mur (2012). [See Supplement 4 for clip of an example observer partaking in the experiments and Methods for details].
Figure S2. Hierarchical category identification tree. See Methods for details.
Figure S3. Identification responses for less ambiguous stimuli at small apertures. Frequency of stimulus labels in the minor contextual effects group for the Large Color (top row), Small Color (2nd row), and Large Grey (3rd row) conditions. A stimulus frame from each movie clip (clip number shown above image) is shown in the bottom row. Ground truth response is the mode response in the Large Color condition. Most observer responses in the Small Color condition are the same as the ground truth label.
Figure S4. Identification responses for highly ambiguous stimuli at small apertures.
Frequency of stimulus labels in the major contextual effects group for the Large Color (top row), Small Color (2nd row), and Large Grey (3rd row) conditions. A stimulus frame from each movie clip (clip number shown above image) is shown in the bottom row. Ground truth response is the mode response in the Large Color condition. Observer responses in the Small Color condition are highly variable with respect to the ground truth label, and often labelled to be more rigid, “pickupable”, and less penetrable than the ground truth label.
Figure S5: Stimulus arrangements and Representational Dissimilarity Matrices (RDMs) for the pooled data in the Large (A) and small (B) color aperture condition for Dataset 2. For each pair of stimuli, each RDM (right) color codes the dissimilarity. The experimental stimuli have been arranged (on the left) such that their pairwise distances approximately reflect the distances in the RDM (multidimensional scaling, dissimilarity: distances, criterion: metric stress) [See Supplement 5 and Supplement 6 for movie clips]. In each arrangement, dynamic textures placed close together were also arranged this way in the experiment. The correlations between the high dimensional RDMs and the two-dimensional Euclidean distances in the figure are 0.82 (Pearson) and 0.82 (Spearman) for the low context and 0.75 (Pearson) and 0.74 (Spearman) for the high context. The RDMs are separately rank-transformed and scaled into [0,1]. In (A) the large aperture conditions, similar stimuli (depicted in blue on RDM) tended to come from the same material category and have similar strength in their intermolecular forces. Dissimilar stimuli in (A) (depicted in yellow on RDM) tended to come from stimuli with contrasting strength in their intermolecular forces. In (B) the Small Color condition, the RDMs show that stimuli within the same category are sometimes similar (in blue). However, this tends to be weaker than in the Large Color condition and so does the tendency for the largest disparities in dissimilarity to arise from flows with large differences in the strength of the intermolecular forces. The MDS visualization on the left shows that color dominates the Small Color arrangement.
Figure S6: Stimulus arrangements and Representational Dissimilarity Matrices (RDMs) for the pooled data in the Large Grey aperture experiment for (A) Dataset 1 and (B) Dataset 2. For each pair of stimuli, each RDM (right) color codes the dissimilarity. The experimental stimuli have been arranged (on the left) such that their pairwise distances approximately reflect the distances in the RDM (multidimensional scaling, dissimilarity: distances, criterion: metric stress) [See Supplement 7 and Supplement 8 for movie clips]. In each arrangement, dynamic textures placed close together were also arranged this way in the experiment. The correlations between the high dimensional RDMs and the two-dimensional Euclidean distances in the figure are 0.76 (Pearson) and 0.75 (Spearman) for Dataset 1 and 0.81 (Pearson) and 0.80 (Spearman) for Dataset 2. The RDMs are separately rank-transformed and scaled into [0,1]. In the Large Grey conditions, similar stimuli (depicted in blue on RDM) tended to come from the same material category and have similar strength in their intermolecular forces. Dissimilar stimuli in (depicted in yellow on RDM) tended to come from stimuli with contrasting strength in their intermolecular forces.
Figure S7. Model performance for similarity judgments in the large (A) and small (B) color aperture experiment for Dataset 2. The deep dimensions tend to explain the similarity arrangements better in the large aperture condition, while the shallow dimensions, in particular color, tends to better account for the small aperture similarity arrangements. The bar graphs show the correlation between the similarity judgment RDM and each of the feature or model-prediction RDMs. Significant correlations between a feature RDM and the similarity judgment RDM are indicated by an asterisk (stimulus-label randomization test, \( p < 0.05 \) corrected for familywise error). Significant differences between models in how well they can account for the similarity judgments are indicated by the black horizontal lines plotted above the bars (stimulus-bootstrap test, \( p < 0.05 \) corrected for familywise error). Error bars show the standard error of the mean based on bootstrap resampling of the stimulus set. The noise ceiling is indicated by the red and green horizontal bars. The noise ceiling is the expected RDM correlation achieved by the (unknown) true model, given the noise in the data. The red bar represents the high noise ceiling, calculated by taking the correlation between each subject’s RDM and the average of all subject RDMs. The green bar represents the low noise ceiling, calculated by taking the correlation between each subject’s RDM and the average of the RDMs belonging to the remaining subjects. The noise ceiling bars are centered on their mean (computed across subjects) with a width that corresponds to their standard error. All models are based on a weighted combination of features.
Figure S8. Model performance for similarity judgments in the Large Grey condition for Dataset 1 (A) and Dataset 2 (B). The deep dimensions tend to explain the similarity arrangements better than shallow dimensions. The bar graphs show the correlation between the similarity judgment RDM and each of the feature or model-prediction RDMs. Significant correlations between a feature RDM and the similarity judgment RDM are indicated by an asterisk (stimulus-label randomization test, $p<0.05$ corrected for familywise error). Significant differences between models in how well they can account for the similarity judgments are indicated by the black horizontal lines plotted above the bars (stimulus-bootstrap test, $p<0.05$ corrected for familywise error). Error bars show the standard error of the mean based on bootstrap resampling of the stimulus set. The noise ceiling is indicated by the red and green horizontal bars. The noise ceiling is the expected RDM correlation achieved by the (unknown) true model, given the noise in the data. The red bar represents the high noise ceiling, calculated by taking the correlation between each subject’s RDM and the average of all subject RDMs. The green bar represent the low noise ceiling, calculated by taking the correlation between each subject’s RDM and the average of the RDMs belonging to the remaining subjects. The noise ceiling bars are centered on their mean (computed across subjects) with a width that corresponds to their standard error. All models are based on a weighted combination of features.
Figure S9. Relationship of Category features to the large aperture human similarity arrangements. The bar graphs show the correlation between the similarity judgment RDM and each of the category feature RDMs for dataset 1 (top) and 2 (bottom) in the Large Color (left) and Large Grey (right) conditions. The factors best related to human similarity are those that indicate material properties relevant to simulation, such as the strength of the material’s intermolecular bonds or wind. Significant correlations between a feature RDM and the similarity judgment RDM are indicated by an asterisk (stimulus-label randomization test, \( p < 0.05 \) corrected for familywise error). Significant differences between models in how well they can account for the similarity judgments are indicated by the black horizontal lines plotted above the bars (stimulus-bootstrap test, \( p < 0.05 \) corrected for familywise error). Error bars show the standard error of the mean based on bootstrap resampling of the stimulus set.
Figure S10. Relationship of Affordances and Material Attribute features to the large aperture human similarity arrangements. The bar graphs show the correlation between the similarity judgment RDM and each of the action and attribute feature RDMs for dataset 1 (top) and 2 (bottom) in the Large Color (left) and Large Grey (right) conditions. The attributes that best correlate with human similarity are pickupability, penetrability, and rigidity which suggests a role of our action decisions in the depth of the visual system’s representation of dynamic flow. Significant correlations between a feature RDM and the similarity judgment RDM are indicated by an asterisk (stimulus-label randomization test, $p<0.05$ corrected for familywise error). Significant differences between models in how well they can account for the similarity judgments are indicated by the black horizontal lines plotted above the bars (stimulus-bootstrap test, $p<0.05$ corrected for familywise error). Error bars show the standard error of the mean based on bootstrap resampling of the stimulus set.