Perceiving animacy from shape

Filipp Schmidt Department of Psychology, Justus-Liebig-University Giessen, Giessen, Germany

Mathias Hegele Department of Psychology, Justus-Liebig-University Giessen, Giessen, Germany

Roland W. Fleming Department of Psychology, Justus-Liebig-University Giessen, Giessen, Germany

Superordinate visual classification—for example, identifying an image as "animal," "plant," or "mineral"—is computationally challenging because radically different items (e.g., "octopus," "dog") must be grouped into a common class ("animal"). It is plausible that learning superordinate categories teaches us not only the membership of particular (familiar) items, but also general features that are shared across class members, aiding us in classifying novel (unfamiliar) items. Here, we investigated visual shape features associated with animate and inanimate classes. One group of participants viewed images of 75 unfamiliar and atypical items and provided separate ratings of how much each image looked like an animal, plant, and mineral. Results show systematic tradeoffs between the ratings, indicating a class-like organization of items. A second group rated each image in terms of 22 midlevel shape features (e.g., "symmetrical," "curved"). The results confirm that superordinate classes are associated with particular shape features (e.g., "animals" generally have high "symmetry" ratings). Moreover, linear discriminant analysis based on the 22-D feature vectors predicts the perceived classes approximately as well as the ground truth classification. This suggests that a generic set of midlevel visual shape features forms the basis for superordinate classification of novel objects along the animacy continuum.

Introduction

Categorization is the process by which visual and cognitive systems organize similar signals into a common class so that knowledge about one item can be generalized to others. Without categorization, the visual system would be overwhelmed by the infinite number of unique items that could be discriminated from one another but whose characteristics must be inferred for each item de novo.

Object categorization can occur at different levels of abstraction: on a superordinate, basic, or subordinate level (e.g., Rosch, 1978/1999). Categorization at the basic level maximizes within-category similarity relative to between-category similarity (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). In other words, members of the same basic object category share most common features when compared with members of other basic categories (e.g., bird vs. tree). Members of the more abstract superordinate categories share fewer features (e.g., animal vs. plant) and members of the more specific subordinate category share more features but also share features with members of other subordinate categories (e.g., fox sparrow vs. house sparrow).

One of the most important distinctions is the superordinate categorization between animate and inanimate objects: It arises early in infancy (Opfer & Gelman, 2011) and presumably has evolved to maximize detection of predators and potential food sources (Barrett, 2005). Although superordinate categorizations are relatively difficult because they involve organization using only few and abstract features, there is evidence that humans can distinguish animate and inanimate objects very well. For example, in speeded-categorization tasks with limited exposure duration, categorization of animacy (i.e., the detection of animals) is easier and faster than basic-level categorization (e.g., Macé, Joubert, Nespoulous, & Fabre-Thorpe, 2009; Praß, Grimsen, König, & Fahle, 2013; Wu, Crouzet, Thorpe, & Fabre-Thorpe, 2014). This suggests there might be dedicated neural hardware to distinguish animals from other objects very quickly, based on early, and relatively coarse, image representations (Cauchoix, Crouzet, Fize, & Serre, 2016; Fabre-Thorpe, 2011; Mack & Palmeri, 2015; and not based on simple low-level image statistics; e.g., Cichy, Pantazis,

Citation: Schmidt, F., Hegele, M., & Fleming, R. W. (2017). Perceiving animacy from shape. *Journal of Vision*, *17*(11):10, 1–15, doi: 10.1167/17.11.10.

doi: 10.1167/17.11.10 Received March 6, 2017; published September 28, 2017 ISSN 1534-7362

& Oliva, 2014; Wichmann, Drewes, Rosas, & Gegenfurtner, 2010). Key features for that rapid classification seem to be the size of the animals relative to the background, whether the animals are in typical postures, and whether distinctive animal features (such as eyes or limbs) are visible (Delorme, Richard, & Fabre-Thorpe, 2010).

When studying categorization in tasks with unlimited or long decision time and exposure duration, where basic-level categorization is superior to superordinate categorization (e.g., Rosch et al., 1976), there is also evidence for a fundamental (superordinate) distinction between animate and inanimate objects in human vision. For example, animacy is the top-level distinction when clustering a large number of object images according to similarity judgments by human observers (Mur et al., 2013). Strikingly, it is also expressed in the neural representation in both human and monkey inferior temporal cortex (IT; Kiani, Esteky, Mirpour, & Tanaka, 2007; Kriegeskorte et al., 2008)—and human response speed in categorization tasks correlates with the distance of an object to category boundaries in this representational space (Carlson, Ritchie, Kriegeskorte, Durvasula, & Ma, 2014). Again, specialized neural hardware appears to be dedicated to achieving this task, although the details are still under debate (Grill-Spector & Weiner, 2014; Proklova, Kaiser, & Peelen, 2016).

Interestingly, it has been shown that perception of animacy also arises from highly simplified dynamic stimuli. First, simple geometric moving shapes can be perceived as animate agents (e.g., Heider & Simmel, 1944; Scholl & Tremoulet, 2000), given specific motion cues (for a review, see Scholl & Gao, 2013). For example, darts might be perceived as "chasing" a user-controlled disc when they are automatically oriented toward ("facing") it—while actually moving randomly across the display (Gao, McCarthy, & Scholl, 2010). Second, animacy can almost instantaneously be perceived from point–light displays that represent, for example, moving humans by light points at the positions of their joints (Johansson, 1973). Here, the animate agents are not defined by a single moving shape but by the combined motion of these light points, which can be interpreted as cues to the invisible global shape of a moving agent. In line with this, observers rate point–light displays of different animals higher on perceived animacy compared with scrambled or inverted versions of these displays (Chang & Troje, 2008; Troje & Chang, 2013). Animacy can also be perceived from point–light displays of artificial "creatures" with rod-like limbs that artificially evolved for performing locomotion—and whose complex articulated motion is different from the familiar motion of animals (Pyles, Garcia, Hoffman, & Grossman, 2007).

Both perceived animacy from simple geometric moving shapes and from point–light displays show that there clearly are motion cues that observers use to make the superordinate distinction between animate and inanimate objects (see Kawabe, 2017). But interestingly, very few studies investigated shape cues to animacy (e.g., Delorme et al., 2010; Jozwik, Kriegeskorte, & Mur, 2016; Wilder, Feldman, & Singh, 2011) although shape is known to be the most relevant feature in visual object categorization (Biederman, 1987; Logothetis & Sheinberg, 1996).

Jozwik et al. (2016) tested the explanatory power of a feature-based model for similarity judgments (and for neural representation) that also included a number of shape features. As similarity judgments typically produce a fundamental distinction between animate and inanimate objects, the authors implicitly tested the role of shape features for animacy categorizations. The model contained 120 feature dimensions that were generated by human observers and included shape features (e.g., curved, cylindrical, spiky), but also object parts (e.g., horns, wings, hand/fingers; about two-thirds of the dimensions), objects (e.g., tree, waterfall, leaves), colors (e.g., green, brown, gray), materials/textures (e.g., fur, metallic, wet/water), and others. On a set of 96 images of familiar objects, the model predicted neural representation in IT with equal accuracy as a categorical labeling model, suggesting that those features "serve as stepping stones toward a representation that emphasizes categorical boundaries or higher-level semantic dimensions" (Jozwik et al., 2016, p. 218)—in accordance with the notion of specialized feature detectors optimized for category discrimination (Sigala & Logothetis, 2002; Ullman, 2007). At the same time, the model performed significantly worse than the categorical model when predicting the full range of behavioral similarity ratings for the same images—but could still to some extent predict the animate–inanimate distinction. By analyzing the dimensional weights, Jozwik et al. (2016) identified those model features that contributed the most to the similarity ratings—most prominent was *head*, followed by the colors *red* and *green*, and the shape features *symmetrical* and *coiled*. This shows that in a set of familiar objects, similarity ratings that produce a categorization between animate and inanimate objects are not as much based on basic shape cues as on higher level, more semantic representations.

A more parametric approach for identifying basic shape cues for superordinate categorization was established by Wilder et al. (2011). The authors argued that categorization might be driven by differences in the global part structure between objects. To retrieve this part structure, they analyzed the medial axis representation of shapes—a skeletal representation by local symmetry axes derived from the shape boundaries (e.g., Blum, 1973; Kovacs, Feher, & Julesz, 1998; Siddiqi, Shokoufandeh, Dickinson, & Zucker, 1999).

Specifically, they built maximum a posteriori (MAP) skeletons (Feldman & Singh, 2006) which produce one skeletal axis per part for large sets of natural two-dimensional (2-D) shapes of animals and leaves. Then, they constructed a number of Bayes classifiers using different features derived from these skeletal representations, and identified the one classifier that distinguishes best between the two sets of skeletons (animals vs. leaves). This "best" model used only two features: (1) the number of skeletal parts and (2) the total signed axis turning angle, thereby effectively distinguishing an animal stereotype with multiple, curvy limbs, from a leaf stereotype with fewer, straighter limbs. The authors asked participants to classify linear morphs between the animal and leaf silhouettes, thereby limiting the influence of semantic knowledge on participants' judgments. The resulting classification probabilities for these novel stimuli were very well explained by the Bayes classifier (trained on the original shapes). After showing that this fit was superior to that of a number of plausible alternative models, Wilder et al. (2011) concluded that at least some aspects of superordinate categorization could be understood as a simple statistical decision reflecting skeletal properties of the shapes (here: number and curvature of limbs).

Thus, it seems that the visual system uses midlevel shape features to classify objects into superordinate categories, presumably as a result of higher-level semantic organization acquired by lifelong learning processes. First, we learn which objects belong to which class (and to what extent they are typical). Based on this higher-level semantic organization, our visual system derives statistical shape features from the known exemplars in each class (weighted by their typicality). Finally, we can use these features to classify novel objects into previously established categories.

In the present study, we set out to identify shape features that define superordinate categories of animals, plants, and minerals for unfamiliar and atypical items. With these three categories, we aimed to span the continuum between animate and inanimate with animals as prototypical animate objects, minerals as inanimate objects, and plants in between (typically, plants were previously included in the class of inanimate objects; e.g., Jozwik et al., 2016)—although previous studies used sometimes finer gradations within classes (e.g., Sha et al., 2015, within the class of animals). In two experiments, we collected typicality ratings (e.g., "how much does an object look like an animal?") and ratings on midlevel shape features (e.g., "how much does an object appear elongated?") from different groups of participants for photographs of unfamiliar and atypical members of the three categories. As a result, we can test (1) how much of the variance in participant's responses can be explained by these midlevel shape features and (2) which features contribute the most to the perception as animal, plant, or mineral.

Note that by using unfamiliar shapes, we aim to reduce the influence of higher-level semantic judgments on participants' ratings (see Wilder et al., 2011; in contrast to Jozwik et al., 2016) for two reasons: first, variance in the ratings will be much higher for unfamiliar compared with familiar shapes, giving us more variance to explain; second, unfamiliar shapes will to some extent decouple ground truth class from perceived class, allowing us to investigate the role of shape features distinctively for object *appearance*. Note that by using rich images of textured and colored objects, we aim to identify the role of midlevel shape features derived from higher-level shape representations (see Jozwik et al., 2016; in contrast to Wilder et al., 2011).

Experiment 1 (typicality rating)

Materials and methods

Participants

Thirteen students from the Justus-Liebig-University Giessen, Germany, with normal or corrected vision participated in the experiment for financial compensation. All participants gave informed consent, were debriefed after the experiment, and were treated according to the ethical guidelines of the American Psychological Association. All testing procedures were approved by the ethics board at Justus-Liebig-University Giessen and were carried out in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki).

Stimuli

We performed an extensive internet search to find 25 images for each of the three classes (animals, plants, minerals), which can be expected to be (1) unfamiliar to our sample of undergraduates, and/or (2) atypical for the class of which they are a member. Atypical members of a category are defined by the extent to which they fail to share the (shape) features with the prototypical member (e.g., penguin vs. prototypical bird; Jolicoeur, Gluck, & Kosslyn, 1984). For example, we collected images of rainforest plants and animals as well as images of animals and minerals exhibiting shape features reminiscent of plants. Although we performed no formal test to evaluate unfamiliarity, given that the stimuli were acquired in many cases from specialist sources it is unlikely that a typical participant would be able to name more than a few of them.

All items were cut out from their background. Their final size differed with a maximal height and length of

15.38° and 28.40° of visual angle, respectively. Note that it was not possible to obtain copyright permission for many of the images so that we can only show some of our stimuli here.

Procedure

Stimuli were presented in color on a black background, using a Dell U2412M monitor (Dell Technologies, Round Rock, TX) at a resolution of 1,920 × 1,200 pixels. Participants were seated about 50 cm away from the monitor. They completed three blocks of 75 trials per block (i.e., all images from all classes in every block). In each trial, a single image was presented and participants moved a slider to indicate the extent to which the image "does NOT look like an ANIMAL/a PLANT/a MINERAL" or "looks A LOT like an ANIMAL/a PLANT/a MINERAL" to them. They completed each rating by pressing the spacebar. In each block, participants rated each image with respect to its typicality for one of the three classes. The order of blocks was counterbalanced between participants. Within each block, images were presented in random order. Stimulus presentation was controlled by MATLAB using the Psychophysics Toolbox extension (Kleiner, Brainard, & Pelli, 2007).

Results and discussion

Each participant rated all 75 stimuli three times, for how much they looked like typical animals, plants, and minerals. Note that although it would be desirable to test much larger stimulus sets, our number of stimuli is similar to that of previous studies with real-world images (e.g., Jozwik et al., 2016, used 96 images).

First, we calculated the correlation between the three independent ratings for each individual image and participant (i.e., across all 975 rating triplets) to obtain a measure for stimulus ambiguity. The higher the resulting correlation coefficients, the more ambiguous were our stimuli. The coefficients show that many images were indeed rather ambiguous, otherwise we would have expected stronger negative correlations (animal–plant: $R = 0.15$; 95% CI [0.09, 0.21]; animal–mineral: $R = -0.41$; 95% CI [–0.46, –0.36]; plant-mineral: $R = -0.47$; 95% CI [–0.51, –0.41]).

At the same time, the pattern of mean responses across participants was far more orderly. Figures 1A and B show mean responses across participants in the three-dimensional space spanned by the three independent ratings. Rather than occupying the full space of possible responses, the data fill only a highly constrained subspace: There is a clear and systematic tradeoff between rating objects in the three different scales, such that they fall close to the negative diagonal. In other words, the more a given image tended to look like an animal, the less it tended to look like a plant or mineral, and vice versa. Note that there is nothing about the task that constrains the responses in this way: Thus, it tells us something about the nature of the internal representation, namely that the ratings are likely derived from attribution of items to mutually exclusive classes. This tradeoff was confirmed by a principal component analysis, which revealed that the first two principal components of the mean ratings account for 98% of the variance, indicating a near-planar arrangement. The mean data are replotted in the 2-D space spanned by the first two principal components (Figure 1C), along with error bars indicating the standard error of the mean and the centroids of the three ground truth classes. Note that the first principal component distinguishes minerals from animals and plants, suggesting a dominance of the animate–inanimate distinction over the distinction between animals and plants.

Inter-rater reliability scores, defined by the average correlation coefficient between image ratings for each possible pair of participants, was $r = 0.61$ across all images ($r = 0.47$, $r = 0.52$, and $r = 0.78$ for the class of animals, plants, and minerals, respectively). The distribution of ratings is distinctly bimodal (Figure 1D), with a strong tendency to give ratings of 0 or 1, and a much weaker tendency to give ratings in between. This indicates that participants tended to respond categorically: A given stimulus generally looked either entirely like an animal or not at all like an animal (and likewise for plants and minerals). Thus, the broad distribution of mean responses reflects the fact that a given stimulus was seen as belonging to different classes by different participants (rather than a direct reporting of uncertainty experienced by individual participants). To make sure that this finding does not result from participants' effort to provide consistent judgments across all three ratings (e.g., by rating an image as low in plant-ness and mineral-ness because they remember having given it a high animal-ness rating in a previous trial), we repeated the experiment with three independent groups of $n = 9$ observers and found strong correlations between the average image ratings of the within-subject and between-subject designs (animal ratings: $R = 0.92$, $p < 0.001$; plant ratings: $R = 0.95$, $p < 0.001$; mineral ratings: $R = 0.94$, $p < 0.001$; overall correlation: $R = 0.94$, $p < 0.001$).

The ambiguity of the stimuli was not accidental. By design, we selected stimuli that were intended to yield false classifications so that we could decouple ground truth class membership from perceived class membership. In Figure 2, we plot the mean ratings across items and participants for each ground truth class, yielding a form of confusion matrix. Cells along the diagonal reflect "correct" responses, while off-diagonal cells indicate the extent to which items were perceived to belong to classes other than the ground truth. The

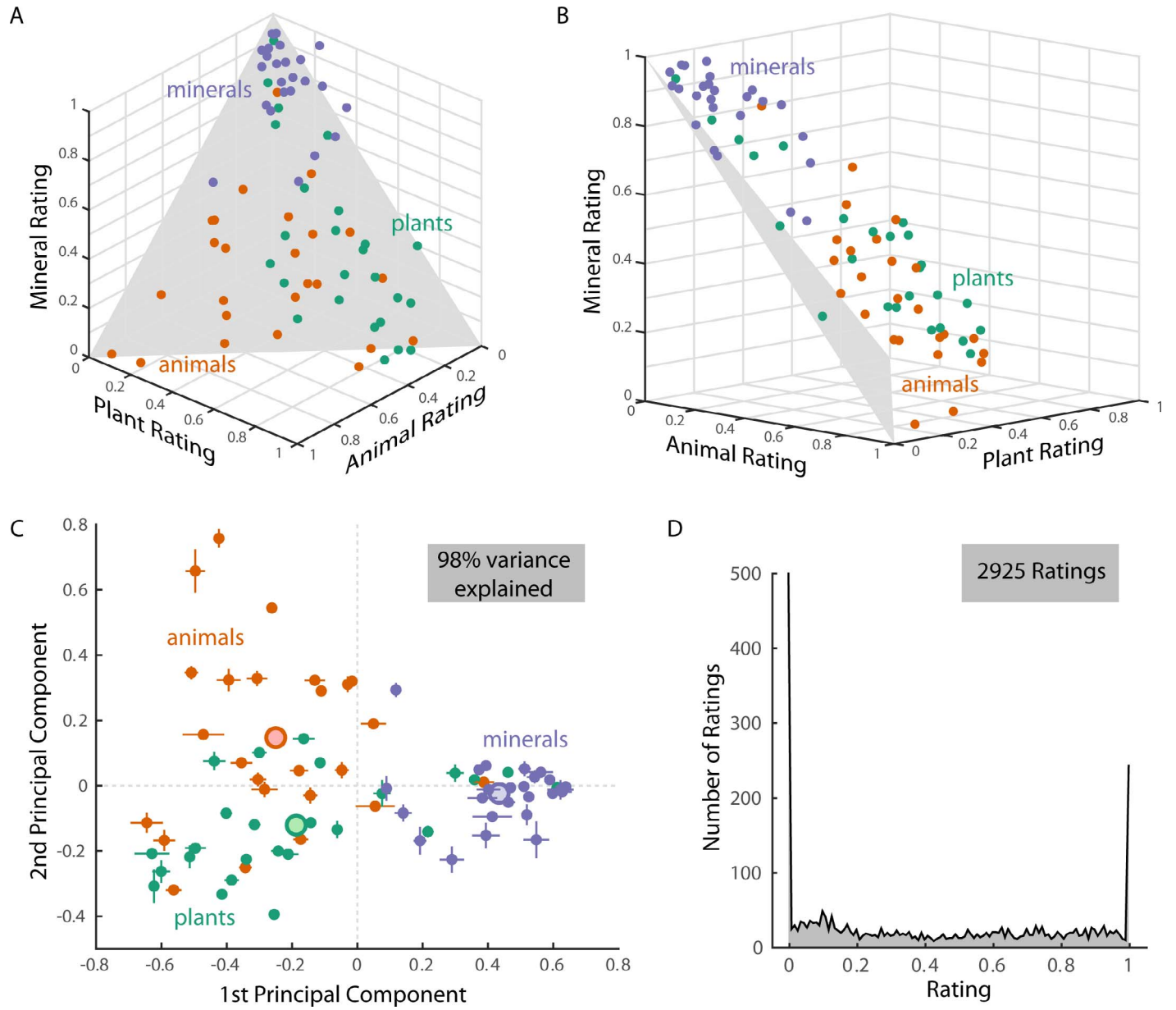

Figure 1. Results of Experiment 1. (A) Points indicate mean ratings, color-coded by ground-truth class. The gray triangle indicates the negative diagonal plane ($x + y + z = 1$). (B) Side view of (A) revealing the constrained nature of the response distribution. (C) Mean ratings replotted in the 2-D space spanned by the first two principal components. Error bars indicate standard errors; large points indicate centroids of each class. (D) Rating distribution across all participants and stimuli.

pattern of responses indicates that while most stimuli were perceived "correctly" (i.e., animals on average tended to receive the high ratings for "looking like an animal"), there were also substantial "errors," in which items belonging to one class were rated as looking most strongly like a member of another class. In each cell, we also show example stimuli with corresponding responses, and in most cases both diagonal and off-diagonal items make intuitive sense. The question we sought to answer in the second experiment was the extent to which midlevel shape features could predict the pattern of responses observed in Experiment 1.

Experiment 2 (shape feature rating)

Materials and Methods

Participants

Twenty students from the Justus-Liebig-University Giessen, Germany, with normal or corrected vision participated in the experiment for financial compensation. No participant took part in Experiment 1. Participants were treated as in Experiment 1.

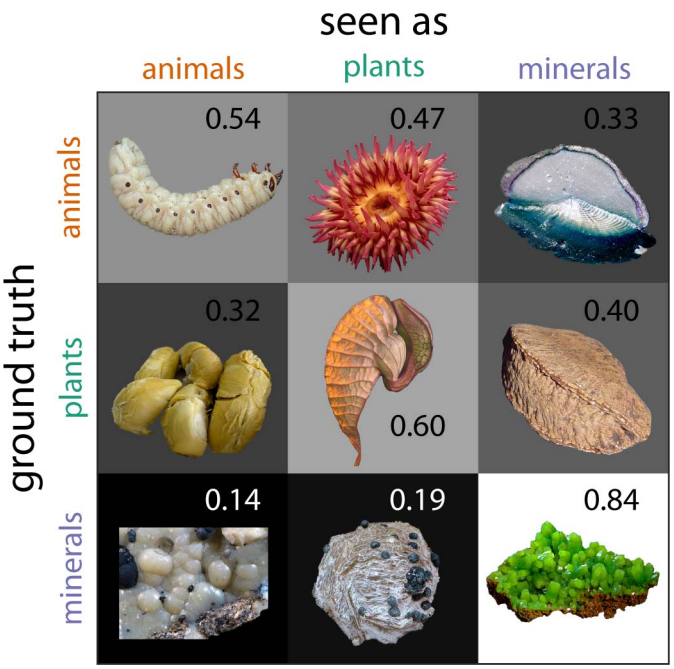

Figure 2. Mean ratings of Experiment 1 per ground truth class (rows) and perceived class (columns). Each cell gives the mean rating, with darker background indicating lower ratings, and shows an example from the ground truth class with relatively high ratings in the respective perceived class. All images are reprinted with permission (from left to right, from top to bottom: "Goliath beetle larva" by Petr Mückstein, 2014; "Sea anemone" by George Grall, 2010; "By-the-wind-sailor" by Natalie Wetherington, 2010; "Durian" by Travis Kaya, 2011; "Aristolochia grandiflora" by Ruth King, 2008; "Brazil nut" by Horst Frank, 2004; "Apatite" by Tom Loomis [Dakota Matrix Minerals]; "Apatite with Cookeite" by Tom Loomis [Dakota Matrix Minerals]; and "Pyromorphite" by Anton Watzl).

Stimuli

Stimuli were the same as in Experiment 1.

Midlevel shape features

Candidate midlevel shape features for animacy (Table 1) were either taken from the literature (e.g., symmetrical and has bumps, Jozwik et al., 2016; number of parts; Wilder et al., 2011), from our previous work (e.g., appears elongated, van Assen, Barla, & Fleming, 2016), and from a brainstorming session within our research group (all were naive to the results of Experiment 1). For rating purposes, the shape features were either formulated as opposites or as statements, depending on which of both was more intuitive when phrased in German.

Procedure

To keep the length of experimental sessions within reasonable bounds, participants were split into two groups: 10 participants rated the shapes with respect to the 11 *opposites* and 10 participants with respect to the 11 *statements*. Before the start of the experiment, each participant was presented with a subset of the stimuli and with an overview of the 11 rating scales, to provide information about the range of different images and ratings that had to be expected.

The experimental setup was the same as in Experiment 1. Participants completed 825 trials. In each trial, a single image was presented together with a single rating scale and participants moved a slider to indicate the extent to which the image was in accordance (1) with one of the two opposites (*opposites*), or (2) with the displayed statement from "true" to "false" or from "many/high" to "none" (*statements*). They completed their rating by pressing spacebar. Images were presented in random order. Stimulus presentation was controlled by MATLAB using the Psychophysics Toolbox extension (Kleiner et al., 2007).

Results and discussion

Mean ratings per feature per image are shown in Figure 3, sorted into the three ground truth categories, and within each ground truth category in ascending order of the typicality ratings from Experiment 1. The shape features (rows) are clustered by their similarity, derived from the responses across all stimuli in Experiment 2–with the clustering solution based on squared Euclidean distances (Ward's minimum variance method) and yielding a reasonable fit of the data (cophenetic correlation: $r = 0.79$). The lack of strongly pronounced vertical block–structure in the sorted matrices shows that the different features are not highly redundant with one another, but instead represent distinct, independent shape features. This lack of compelling cluster structure potentially suggests that the visual system uses a complex, high-dimensional combination of midlevel shape features to infer superordinate categorization. Inter-rater reliability across all features was $r = 0.35$, and varied between $r = 0.17$ and $r = 0.65$ for the different features, suggesting that verbal labels for shape features are quite ambiguous, or that judging such features in complex, natural images is a hard task for naive observers. Also, inter-rater reliability was different between both participant groups (opposites: $r = 0.42$; statements: $r = 0.29$); however, it is not clear whether this difference is explained by the differences between participants, shape features, or formulation as opposites or statements.

To evaluate the role of each feature in the image ratings obtained in Experiment 1, we looked at the raw correlations between the 22 feature ratings and animal, plant, and mineral ratings, respectively. In Figure 4, we

Feature labels		Description
Opposites		
1. rounded	*'rounded'*	*'angular'*
2. symmetrical	*'symmetrical'*	*'asymmetrical'*
3. rough	*'rough/spiky'*	*'smooth'*
4. simple	*'simple'*	*'complex'*
5. systematic	*'regular/systematic/orderly'*	*'random/chaotic/disorganized'*
6. repetitive	*'consists of reoccurring parts'*	*'consists of unique parts'*
7. curved	*'bent/curvy parts'*	*'straight/linear parts'*
8. pointed	*'pointed part'*	*'rounded off parts'*
9. single part	*'consists of one parts'*	*'consists of multiple parts'*
10. single texture	*'homogeneous texture'*	*'has multiple textures'*
11. bulky	*'fat/bulky shape or parts'*	*'thin/wispy shape or parts'*
Statements		
1. front/back	*'front and back (orientation) noticeable'*	
2. branched	*'has a branched structure'*	
3. concavities	*'has concavities/holes/lumps'*	
4. geometric	*'similarity to simple geometric shapes'*	
5. elongated whole	*'appears elongated'*	
6. tapered	*'tapered towards one or both ends'*	
7. elongated parts	*'has elongated parts'*	
8. main part	*'has one obvious main part'*	
9. flat	*'is flat'*	
10. protuberances	*'has eversions'*	
11. bumpy	*'has bumps'*	

Table 1. Feature labels and descriptions of midlevel shape features (see Table A1 in the Appendix for the original German descriptions).

show the top three features with the strongest positive correlations with the ratings of animal, plant, and mineral.

However, these raw correlations do not consider to what extent each feature is explaining shared variance of animal, plant, and mineral ratings. Therefore, we performed a PCA, including all 22 features. We found that there is no single feature that explains a majority of the variance but rather many features that each explain some part of the variance (Figure 5A): Only when including 10 principal components, it is possible to explain $> 90\%$ of the variance in the ratings. This further confirms the lack of redundancy between shape features, indicating that the space of midlevel features is high dimensional. In Figure 5B–D, we show word clouds for each of the three classes illustrating the relative contribution of each feature based on their regression weights.

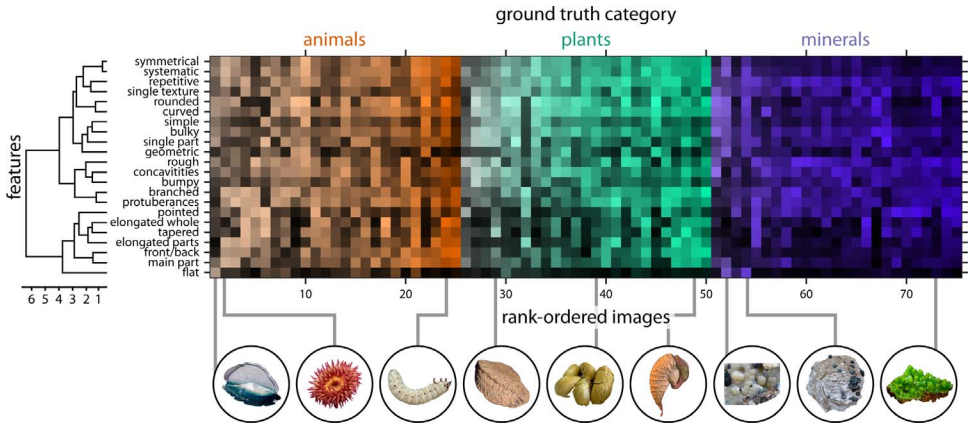

Figure 3. Results of Experiment 2. Matrix of mean ratings organized by features (rows) and images (columns). Features are sorted according to their similarity (see dendrogram). Images are rank-ordered by their ratings within each ground-truth class. Colors indicate ground-truth class of animals, plants, and minerals; saturation indicates how much each image is seen as animal, plant, or mineral; intensity indicates how much each feature is rated for the respective image.

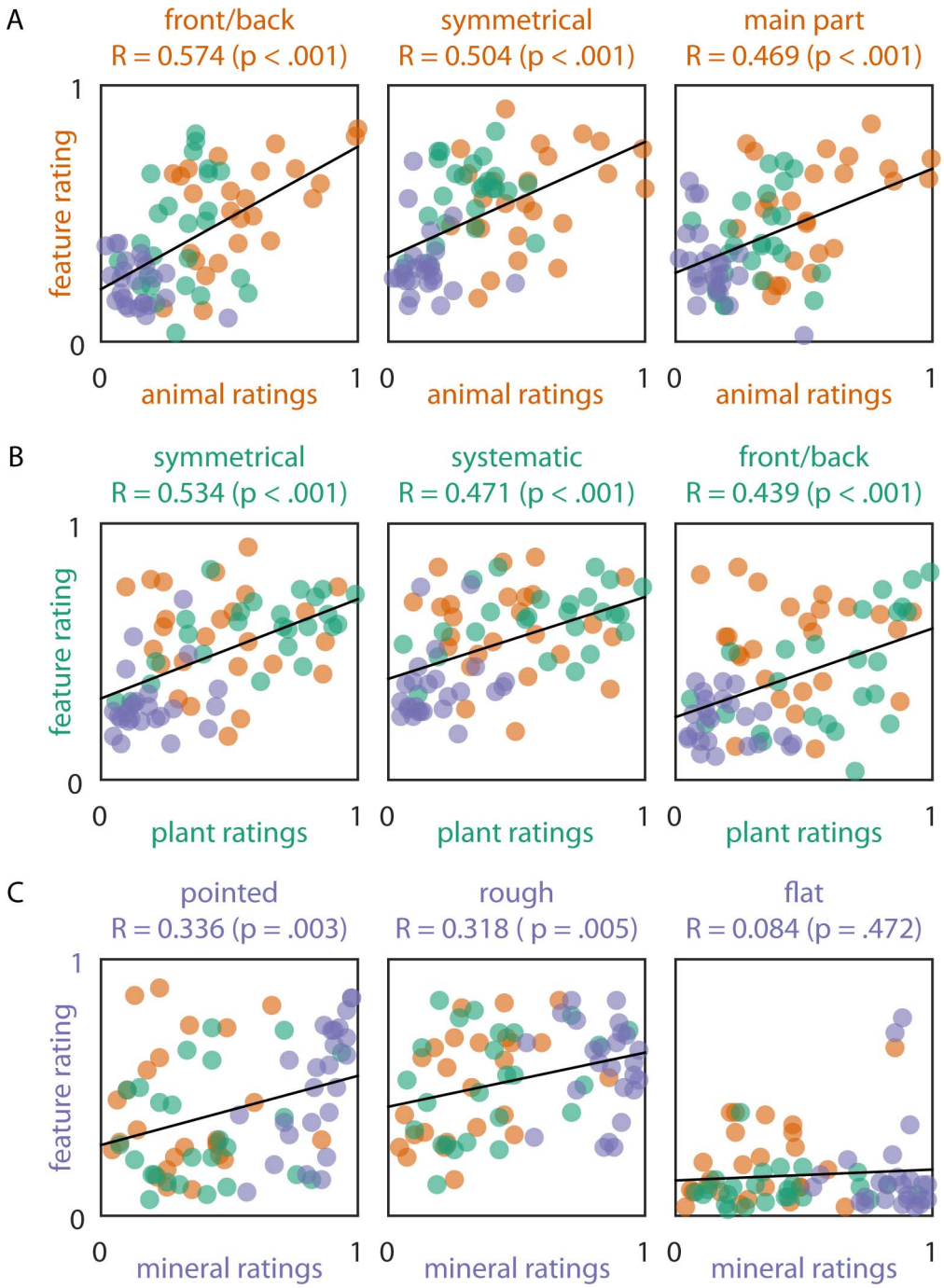

Figure 4. Top three features that were correlated with ratings of (A) animal, (B) plant, and (C) mineral, respectively. Each panel shows the correlation between the ratings of a single feature with the class ratings for each image (discs). Discs are colored according to the ground truth classification of the image. The title of each panel shows the feature label, the correlation coefficient, and the *p* value.

Though this analysis provides significantly more information compared with the raw correlations, it is still not conclusive about which features best *distinguish* between the classes. To evaluate this, we calculated receiver operating characteristic (ROC) curves for multilabel classifications and for each class identify the area under the curve (AUC) of each feature when distinguishing between the respective class and the other two classes combined (Figure 6A). This allows to identify those single features that do significantly contribute to these distinctions. For example, for this set of stimuli, items that look like animals (rather than plants or minerals) tend to have *one main part*, are *symmetrical*, with a *clear front and back* and a *systematic* shape, and tend not to be *rough*. These results hardly change when removing images (Figure

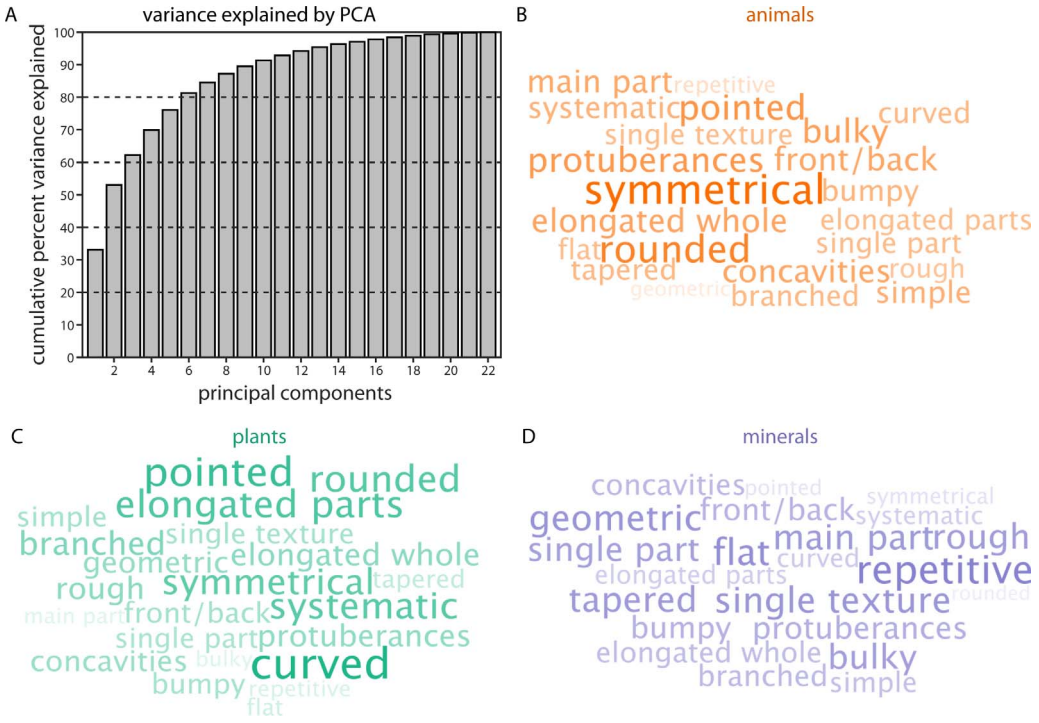

Figure 5. (A) Proportion of variance explained as a function of the number of principal components. (B) Word cloud for the relative contribution of each feature in the perception of animals, (C) plants, and (D) minerals; size and saturation of feature labels is based on the regression weights within the animal–mineral–plant rating space.

6B), suggesting that at least for this set of stimuli the findings are stable.

Finally, to test whether the feature ratings from Experiment 2 could predict the perceived class membership from Experiment 1, we first assigned each item in Experiment 1 to the class for which it received the highest mean rating. Then, we created a linear discriminant analysis classifier to identify three predicted classes. Correspondence between predicted and perceived classes was established using the frequency of co-membership of different items (i.e., the cluster with the most animal items was compared with the "animal" class, and so on), the prediction error was about 15% (i.e., 11 images).

Figure 7 shows for each image the ground truth class, the perceived class, and the predicted class based on the linear discriminant analysis. Interestingly, this analysis based on shape feature ratings predicts the perceived classes approximately as well as the ground truth classification, as shown by the mean correlations between ground truth, perceived, and predicted classes (Figure 8). There was a significant correlation between predicted and perceived class ($R = 0.87$; 95% CI [0.80, 0.92]) that is as strong as that between ground truth and perceived class ($R = 0.70$; 95% CI [0.57, 0.80]) ($z = 1.55$, $p = 0.122$).

General discussion

Shape is the most important cue for visual object recognition and classification; and one of the most fundamental distinctions in classification refers to the distinction between animate and inanimate objects. We know that objects' shapes result from complex interactions of a multitude of forces (e.g., mechanical, chemical, genetic; Ball, 2009) as well as selective pressures that led to the expression of distinctive shape features in animals and plants. This prompts the question of which specific attributes or features humans use to distinguish between animate and inanimate objects—a question that has been addressed by only very few studies to date (Delorme et al., 2010; Jozwik et al., 2016; Wilder et al., 2011). As these studies either included higher level semantic features (such as head, eyes, or limbs; Delorme et al., 2010; Jozwik et al., 2016) or shape cues derived from abstract, global shape representations (Wilder et al., 2011), we specifically tested for the role of midlevel shape features which are represented in-between the levels of semantic knowledge (i.e., object recognition) and global shape. We used colored and textured images of animal, plant, or mineral objects with unfamiliar and atypical shapes, and compared ratings of class membership (Experiment 1) to ratings of midlevel shape features (Experiment 2)

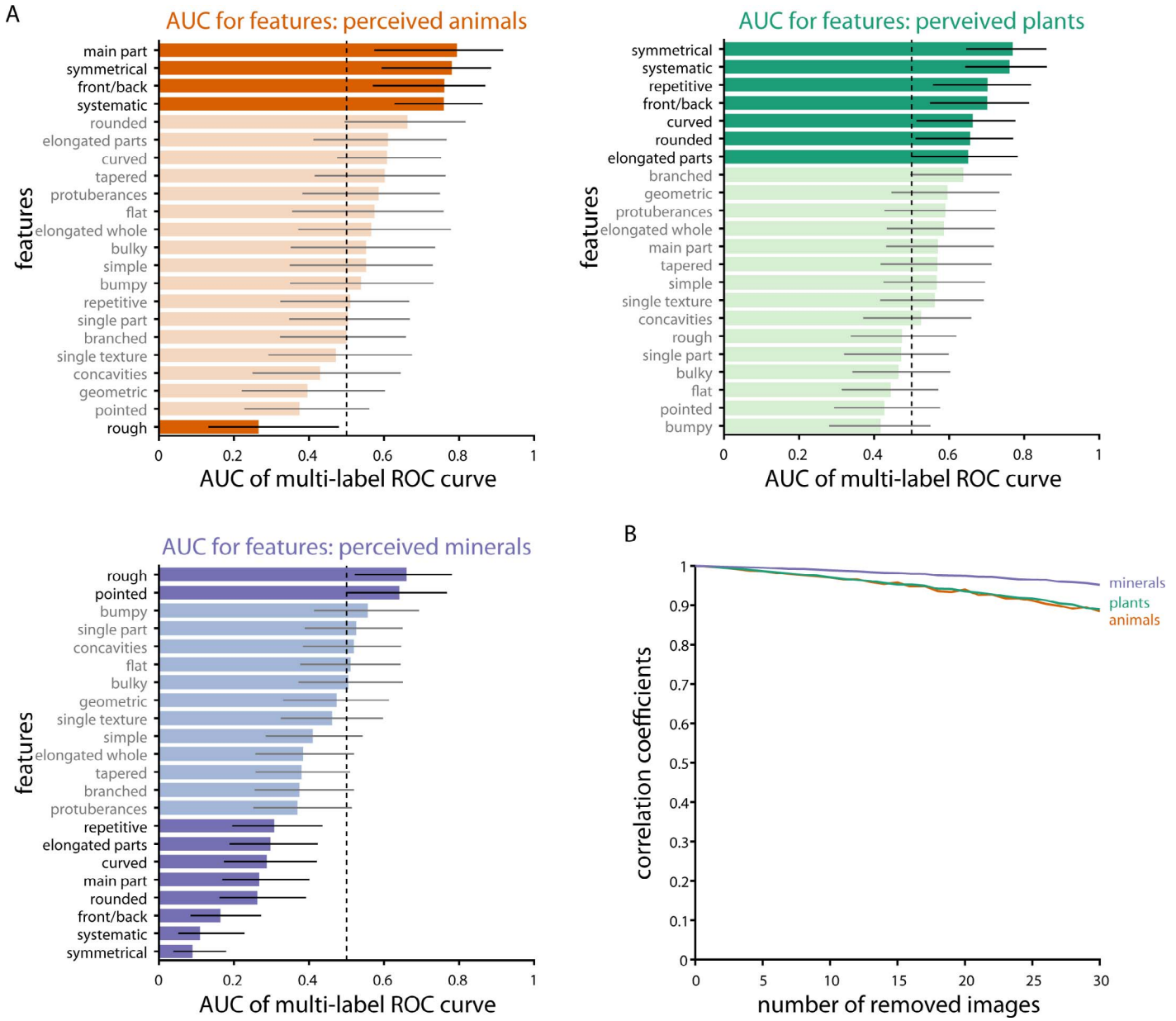

Figure 6. (A) Area under the curve (AUC) of multiclass receiver operating characteristic (ROC) curves for each class (animals, plants, minerals). Features are rank-ordered according to their AUC values; AUC values > 0.5 indicate positive predictors of the class compared with the other two classes, AUC values < 0.5 indicate negative predictors of the class. Error bars are 95% confidence intervals obtained by bootstrapping; features not significantly different from 0.5 are grayed out. (B) Average correlation coefficients between the AUC values from (A) and the AUC values that are obtained when removing n random images from the image set before calculating multiclass ROC curves. The correlation coefficients are calculated separately for each class and plotted as a function of images removed. Notably, even with $n = 30$ images removed, correlation coefficients are still about $r = 0.9$.

obtained from different groups of observers. We found that the classification (rating) as animal, plant, or mineral could be predicted to some extent by the amount to which the depicted objects exhibited particular midlevel shape features.

Specifically, the most prominent features in the objects perceived as animals were *symmetry*, *roundedness*, and *pointedness*, in perceived plants they were *curvedness*, *pointedness*, and *elongated parts*, and in perceived minerals they were *repetitiveness*, *flatness*, and *geometrical shape* (Figure 5B–D). By identifying those features that were most shared by members of one class and most discriminative against the members of the other two classes, we obtain positive and negative shape predictors for all three classes (Figure 6A). For perceived animals, we identified showing a *main part*, being *symmetrical*, showing a *front and back*, and being *systematic* as positive, and being *rough* as

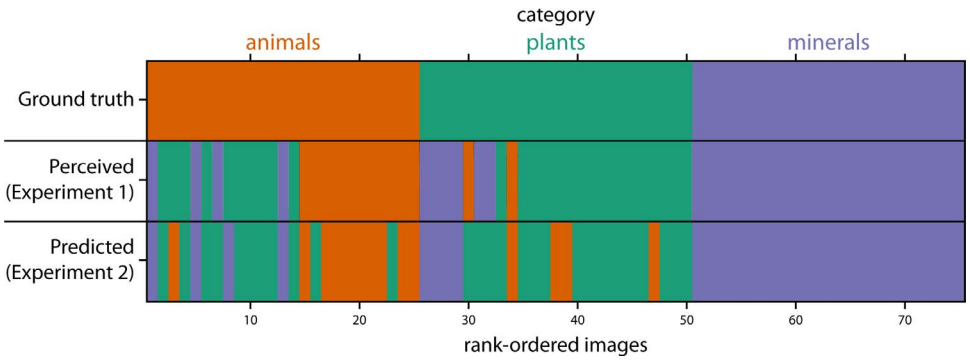

Figure 7. Matrix of membership to ground truth, perceived, and predicted classes (rows) per image (columns). Predicted class membership is derived from linear discriminant analysis based on shape features. Images are rank-ordered by their ratings within each class (as in Figure 3). Colors indicate whether the image is ordered into the animal, plant, or mineral class.

negative shape predictors. For perceived plants, it was being *symmetrical*, *systematic*, and *repetitive*, showing a *front and back*, being *curved* and *round*, and having *elongated parts* as positive shape predictors (with no negative predictors). For perceived minerals, it was being *rough* and *pointed* as positive predictors versus being *repetitive*, having *elongated parts*, being *curved*, showing a *main part*, being *round*, showing a *front and back*, being *systematic* and *symmetrical* as negative shape predictors.

Thus, we identified a set of midlevel shape features that contribute to the fundamental distinction between animals, plants, and minerals. This contribution seems to be relatively independent of higher semantic knowledge or global shape characteristics; for example, midlevel shape features that could be considered as approximations of higher level semantic features for animal-ness (e.g., has one obvious main part ~ headedness; protrusions ~ limbs) did not explain much variance. A similar point can be made with respect to global shape characteristics: Although Wilder et al. (2011) identified global shape cues that could well distinguish between animals and leaf contours (i.e., number and curvature of limbs), such global cues cannot well distinguish between classes within a much richer and more heterogeneous stimulus set as ours. In contrast, we find that although some shape features are more important and distinctive compared with others, most of the variance is explained by many features, each of which contributes a little (see Figure 5A).

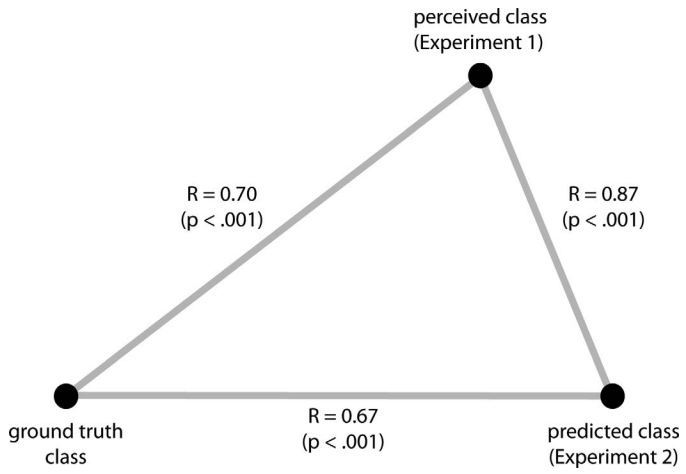

Figure 8. Mean correlations between ground truth, perceived, and predicted classes.

We do not seek to draw the conclusion that the shape features investigated here are in any sense a definitive set: There are likely to be many other midlevel features that contribute to animacy perception, and indeed it may be possible to generate stimuli that appear like animals, plants, or minerals despite having different values for the set of features we tested. The conclusion we draw here is not so much the importance of these specific shape features, but rather the general principle that the visual system can estimate the animacy of unfamiliar items through specific combinations of midlevel features.

Overall, we find that (1) animals are more often perceived as plants than as minerals, (2) plants are sometimes perceived as minerals, sometimes as animals, and (3) minerals are always perceived as minerals. Moreover, shape feature ratings are more similar between perceived animals and plants ($r = 0.88$) compared with between animals and minerals ($r = 0.30$), whereas the similarity between shape feature ratings for plants and minerals is somewhere in between ($r = 0.59$). This supports our assumption that animals, plants and minerals are spread out along an animacy continuum, and that their position along this continuum is reflected in their shape features. Note that this might also reflect the fact that an animal or plant is more of a distinct entity (cf. count noun) and less of a "material" (cf. non-count noun) compared with a mineral—which is often part of a larger (e.g., rock) formation. We suggest that these class-defining properties contribute to the distinction between animals, plants, and minerals via midlevel shape features related

to these properties. For example, our finding that a distinctive feature of animals (vs. minerals and plants) is having an obvious main part (Figure 6A) is incompatible with being a homogeneous material.

More generally, the more distinctive shape features for animal are present in a given object, the more likely it will be identified as animal. Indeed, response times as well as neural representation vary with the amount of perceived animacy on the animacy continuum (e.g., Sha et al., 2015). In more general terms, the time necessary to categorize an object according to some representational boundary follows from the distance of the object representation from this boundary (Carlson et al., 2014). Here, we identified the midlevel shape features that could form the basis of this representational shape space for distinctions between animals, plants, and minerals for unfamiliar and atypical class members.

Note that by using unfamiliar shapes we aimed at identifying features that were independent of semantic knowledge. This is an important difference between our approach and previous studies that tested the role of shape features of intermediate complexity for categorization (e.g., Sigala & Logothetis, 2002; Ullman, 2007; Ullman, Vidal-Naquet, & Sali, 2002)—in these studies, the distinctive features are typically very specific and closely tied to existing semantic labels (e.g., eyes for face class, tires for car class). An exception is a recent study by Long, Störmer, and Alvarez (2017) using unfamiliar "texforms" that are based on animal and object images but only preserve some of the original texture and shape information. The authors find that classification of these "texforms" as animate or inanimate depended on their perceived curvature. Here, we identified more generic shape features that could also form the basis for other classification problems than animacy perception.

As in all previous studies on this topic, the generality of our findings is limited by the selected stimulus images and the initial choice of potentially relevant midlevel shape features. Here, we aimed to provide a proof of concept by showing that a relatively small number of shape features can be used to predict superordinate classification of complex, natural images quite well. Having demonstrated the plausibility of such an approach, further work should use a much larger set of items to seek low-dimensional manifold structure in the feature space, which may suggest novel features that remove redundancies and therefore represent the shape space more compactly. Although technically challenging, another important direction for future work is to synthesize novel images with parametric variations of the putative features, to identify which ones play a causal role in participants' judgments.

Keywords: shape, classification, shape representation, perceptual organization

Acknowledgments

This research was funded by the DFG funded Collaborative Research Center "Cardinal Mechanisms of Perception" (SFB-TRR 135) and the ERC Consolidator award "SHAPE" (ERC-CoG-2015-682859). The authors thank Judith-Larissa Orzschig for image compilation, stimulus preprocessing, and data collection. Raw data from both experiments are available for download at https://doi.org/10.5281/zenodo.848210.

Commercial relationships: none.
Corresponding author: Filipp Schmidt.
Email: Filipp.Schmidt@psychol.uni-giessen.de.
Address: Department of Psychology, Justus-Liebig-University Giessen, Giessen, Germany.

References

Ball, P. (2009). *Shapes: nature's patterns: a tapestry in three parts*. Oxford, UK: Oxford University Press.

Barrett, H. C. (2005). Adaptations to predators and prey. In D. Buss (Ed.), *The Handbook of Evolutionary Psychology* (pp. 200–223). Hoboken, NJ: John Wiley.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, *94*, 115–147.

Blum, H. (1973). Biological shape and visual science (Part I). *Journal of Theoretical Biology*, *38*, 205–287.

Carlson, T. A., Ritchie, J. B., Kriegeskorte, N., Durvasula, S., & Ma, J. (2014). Reaction time for object categorization is predicted by representational distance. *Journal of Cognitive Neuroscience*, *26*, 132–142.

Cauchoix, M., Crouzet, S. M., Fize, D., & Serre, T. (2016). Fast ventral stream neural activity enables rapid visual categorization. *NeuroImage*, *125*, 280–290.

Chang, D. H., & Troje, N. F. (2008). Perception of animacy and direction from local biological motion signals. *Journal of Vision*, *8*(3):5, 1–10, doi:10.1167/8.3.5. [PubMed] [Article]

Cichy, R. M., Pantazis, D., & Oliva, A. (2014).

Resolving human object recognition in space and time. *Nature Neuroscience*, *17*, 455–462.

Delorme, A., Richard, G., & Fabre-Thorpe, M. (2010). Key visual features for rapid categorization of animals in natural scenes. *Frontiers in Psychology*, *1*, 21.

Fabre-Thorpe, M. (2011). The characteristics and limits of rapid visual categorization. *Frontiers in Psychology*, *2*, 243.

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

Gao, T., McCarthy, G., & Scholl, B. J. (2010). The wolfpack effect: Perception of animacy irresistibly influences interactive behavior. *Psychological Science*, *21*, 1845–1853.

Grill-Spector, K., & Weiner, K. S. (2014). The functional architecture of the ventral temporal cortex and its role in categorization. *Nature Reviews Neuroscience*, *15*, 536–548.

Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *American Journal of Psychology*, *57*, 243–249.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics*, *14*, 201–211.

Jolicoeur P., Gluck M. A., & Kosslyn S. M. (1984) Pictures and names: making the connection. *Cognitive Psychology*, *16*, 243–275.

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.

Kawabe, T. (2017). Perceiving animacy from deformation and translation. *i-Perception*, *8*(3), 1–14.

Kiani, R., Esteky, H., Mirpour, K., & Tanaka, K. (2007). Object category structure in response patterns of neuronal population in monkey inferior temporal cortex. *Journal of Neurophysiology*, *97*, 4296–4309.

Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3? *Perception, 36th ECVP Abstract Supplement*.

Kovacs, I., Feher, A., & Julesz, B. (1998). Medial-point description of shape: A representation for action coding and its psychophysical correlates. *Vision Research*, *38*, 2323–2333.

Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., ... Bandettini, P. A. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron*, *60*, 1126–1141.

Logothetis, N. K., & Sheinberg, D. L. (1996). Visual object recognition. *Annual Review of Neuroscience*, *19*, 577–621.

Long, B., Störmer, V. S., & Alvarez, G. A. (2017). Mid-level perceptual features contain early cues to animacy. *Journal of Vision*, *17*(6):20, 1–20, doi:10.1167/17.6.20. [PubMed] [Article]

Macé, M. J. M., Joubert, O. R., Nespoulous, J. L., & Fabre-Thorpe, M. (2009). The time-course of visual categorizations: You spot the animal faster than the bird. *PloS One*, *4*, e5927.

Mack, M. L., & Palmeri, T. J. (2015). The dynamics of categorization: Unraveling rapid categorization. *Journal of Experimental Psychology: General*, *144*, 551–569.

Mur, M., Meys, M., Bodurka, J., Goebel, R., Bandettini, P. A., & Kriegeskorte, N. (2013). Human object-similarity judgments reflect and transcend the primate-IT object representation. *Frontiers in Psychology*, *4*, 128.

Opfer J. E., & Gelman S. A. (2011). Development of the animate–inanimate distinction. In U. Goswami (Ed.), *The Wiley-Blackwell handbook of childhood cognitive development* (pp. 213–238). Oxford, England: Wiley-Blackwell.

Praß, M., Grimsen, C., König, M., & Fahle, M. (2013) Ultra rapid object categorization: Effects of level, animacy and context. *PloS One*, *8*, e68051.

Proklova, D., Kaiser, D., & Peelen, M. V. (2016). Disentangling representations of object shape and object category in human visual cortex: The animate–inanimate distinction. *Journal of Cognitive Neuroscience, X:Y*, 1–13.

Pyles, J. A., Garcia, J. O., Hoffman, D. D., & Grossman, E. D. (2007). Visual perception and neural correlates of novel 'biological motion'. *Vision Research*, *47*, 2786–2797.

Rosch, E. (1978 /1999). Principles of categorization. In E. Margolis & S. Laurence (Eds.), *Concepts: Core readings* (pp. 189–206). Cambridge, MA: MIT Press.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*, 382–439.

Scholl, B. J., & Gao, T. (2013). Perceiving animacy and intentionality: Visual processing or higher-level judgment? In M. D. Rutherford & V. A. Kuhlmeier (Eds.), *Social perception: Detection and interpretation of animacy, agency, and intention* (pp. 197–230). Cambridge, MA: MIT Press.

Scholl, B. J., & Tremoulet, P. D. (2000). Perceptual causality and animacy. *Trends in Cognitive Sciences*, *4*, 299–309.

Sha, L., Haxby, J. V., Abdi, H., Guntupalli, J. S., Oosterhof, N. N., Halchenko, Y. O., & Connolly, A. C. (2015). The animacy continuum in the human ventral vision pathway. *Journal of Cognitive Neuroscience*, *27*:4, 665–678.

Siddiqi, K., Shokoufandeh, A., Dickinson, S. J., & Zucker, S. W. (1999). Shock graphs and shape matching. *International Journal Of Computer Vision*, *35*, 13–32.

Sigala, N., & Logothetis, N. (2002). Visual categorization shapes feature selectivity in the primate temporal cortex. *Nature*, *415*, 318–320.

Troje, N. F., & Chang, D. H. F. (2013). Shape-independent processes in biological motion perception. In K. L. Johnson, & M. Shiffrar (Eds.) , *People watching: Social, perceptual, and neurophysiological studies of body perception* (pp. 82–100). Oxford, UK: Oxford University Press.

Ullman, S. (2007). Object recognition and segmentation by a fragment-based hierarchy. *Trends in Cognitive Sciences*, *11*, 58–64.

Ullman, S., Vidal-Naquet, M., & Sali, E., (2002). Visual features of intermediate complexity and their use in classification. *Nature Neuroscience*, *5*, 682–687.

van Assen, J. J. R., Barla, P., & Fleming, R. W. (2016). Identifying shape features underlying liquid perception. *Perception, 45 (ECVP Abstract Supplement S2)*, 89–89.

Wichmann, F. A., Drewes, J., Rosas, P., & Gegenfurtner, K. R. (2010). Animal detection in natural scenes: critical features revisited. *Journal of Vision*, *10*(4):6, 1–27, doi:10.1167/10.4.6. [PubMed] [Article]

Wilder, J., Feldman, J., & Singh, M. (2011). Superordinate shape classification using natural shape statistics. *Cognition*, *119*, 325–340.

Wu, C. T., Crouzet, S. M., Thorpe, S. J., & Fabre-Thorpe, M. (2014). At 120 msec you can spot the animal but you don't yet know it's a dog. *Journal of Cognitive Neuroscience*, *27*, 141–149.

Appendix

Feature labels	Description	
Opposites		
1. rounded	*'rund'*	*'kantig'*
2. symmetrical	*'symmetrisch'*	*'asymmetrisch'*
3. rough	*'rau'*	*'glatt'*
4. simple	*'einfach'*	*'komplex'*
5. systematic	*'reguläre/systematische Form'*	*'irreguläre/chaotische Form'*
6. repetitive	*'besteht aus sich wiederholenden Teilen'*	*'besteht aus einzigartigen Teilen'*
7. curved	*'kurvige/gebogene Teile'*	*'gerade/geradlinige Teile'*
8. pointed	*'spitze Teile'*	*'abgerundete Teile'*
9. single part	*'besteht aus einem Teil'*	*'besteht aus mehreren Teilen'*
10. single texture	*'hat eine Textur'*	*'hat mehrere Texturen'*
11. bulky	*'dicke/voluminöse Teile/Form'*	*'dünne/feine Teile/Form'*
Statements		
1. front/back	*'vorne/hinten (Orientierung) erkennbar'*	
2. branched	*'hat eine verzweigte Struktur'*	
3. concavities	*'hat Einwölbungen/Löcher'*	
4. geometric	*'Ähnlichkeit mit simplen geometrischen Figuren/Formen'*	
5. elongated whole	*'erscheint langgestreckt'*	
6. tapered	*'zugespitzt an einem Ende/beiden Enden'*	
7. elongated parts	*'hat langgestreckte Teile'*	
8. main part	*'hat einen offensichtlichen Hauptteil'*	
9. flat	*'ist flach'*	
10. protuberances	*'hat Ausstülpungen'*	
11. bumpy	*'hat Beulen'*	

Table A1. Feature labels and descriptions of midlevel shape features (German).