

# Getting “fumpered”: Classifying objects by what has been done to them

Roland W. Fleming

Justus-Liebig-University Giessen, General Psychology,  
Giessen, Germany



Filipp Schmidt

Justus-Liebig-University Giessen, General Psychology,  
Giessen, Germany



**Every object acquires its shape from some kind of generative process, such as manufacture, biological growth, or self-organization, in response to external forces. Inferring such generative processes from an observed shape is computationally challenging because a given process can lead to radically different shapes, and similar shapes can result from different generative processes. Here, we suggest that in some cases, generative processes endow objects with distinctive statistical features that observers can use to classify objects according to what has been done to them. We found that from the very first trials in an eight-alternative forced-choice classification task, observers were extremely good at classifying unfamiliar objects by the transformations that had shaped them. Further experiments show that the shape features underlying this ability are distinct from Euclidean shape similarity and that observers can separate and voluntarily respond to both aspects of objects. Our findings suggest that perceptual organization processes allow us to identify salient statistical shape features that are diagnostic of generative processes. By so doing, we can classify objects we have never seen before according to the processes that shaped them.**

## Introduction

Every object in our environment has been created or altered by some generative process, such as manufacture, biological growth, or self-organization—a leaf has grown, a dune shaped by wind, a tool cast from iron. Although the notion that we can visually infer these generative processes is long-standing (Arnheim, 1974; Leyton, 1989), it has received relatively little scientific attention (Chen & Scholl, 2016; Pinna, 2010; Pinna & Deiana, 2015; Schmidt & Fleming, 2016; Schmidt & Fleming, 2018; Spröte & Fleming, 2013; Spröte, Schmidt, & Fleming,

2016). Identifying exactly what has been done to an object in its past just from its observed shape is far from trivial. For example, it is practically impossible to infer the exact forces and processes responsible for creating the shape of a crumpled ball of paper. Even in less complex cases, it is computationally challenging to separate features based on their cause because features with different causal origins may be superimposed and distributed across the object. Indeed, a given process can yield very different results when applied to different initial objects, and conversely, two objects that have been subjected to very different transformations may yet retain similar overall shapes.

Despite these challenges, there is potentially a close link between the causal history of objects and their classification. A particular causal history (e.g., “twisting”) can produce similar shape features in different objects (e.g., spiralling features). Here, we suggest that even when we cannot infer the exact causal history of an object, we may still be able to classify them based on the distinctive features that result from the processes. By identifying tell-tale “texture-like” statistical regularities in the shape of novel objects, we can group together objects that have a different overall shape but have been created or affected by similar causal processes.

Consider the object in Figure 1E, whose limbs have been subjected to distinct generative processes. It is subjectively relatively easy to identify the different regions based on their statistical surface properties. This suggests that mid-level shape properties (i.e., the complex relationships between local geometrical properties) could provide the basis for categorization of novel objects, even if we cannot infer a full generative model for the object.

Here, we test whether observers can generalize from such statistical regularities by asking them to group novel objects affected by a set of unfamiliar transfor-

Citation: Fleming, R. W., & Schmidt, F. (2019). Getting “fumpered”: Classifying objects by what has been done to them. *Journal of Vision*, 19(4):15, 1–12, <https://doi.org/10.1167/19.4.15>.



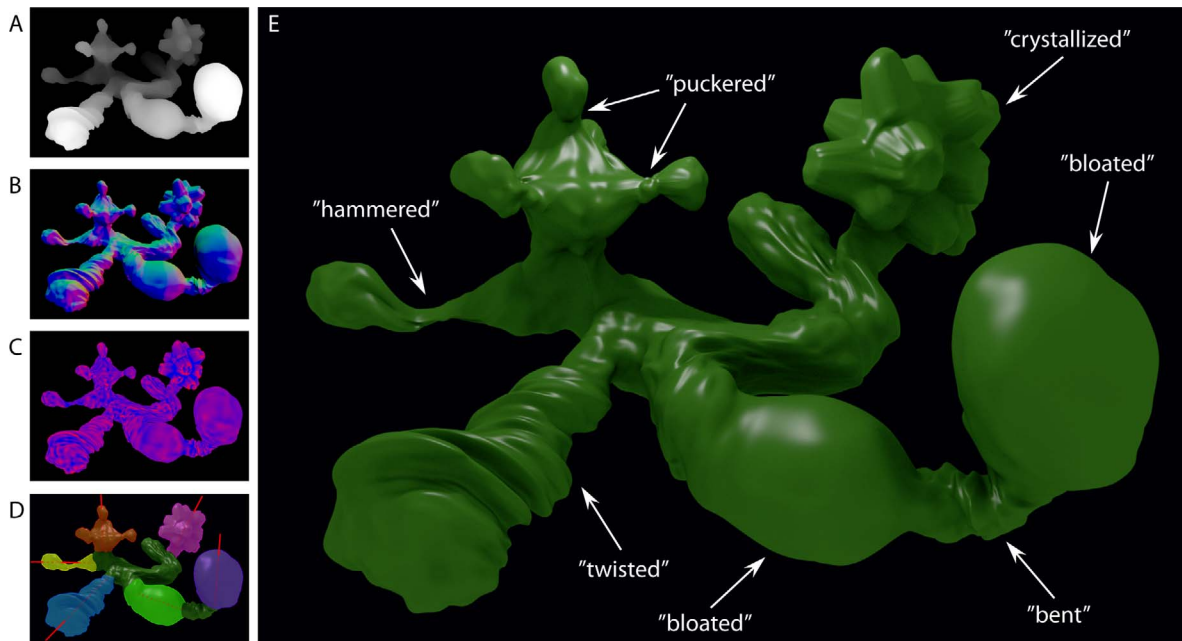


Figure 1. A new look on perceiving shape. Traditional approaches to shape perception involve shape computations based on (A) depths, (B) surface normals, (C) curvatures, or (D) parts. Here, we focus on (E) the causal history of shape, where different objects or different regions of the same object might be distinguished by the generative process that produced their shape (e.g., “twisted”).

mations. If they are able to group objects based on the transformations, this would suggest that they access a representational space containing the relevant telltale features for this classification. We suggest that learning to categorize familiar objects establishes a feature space that provides the basis for similarity judgments and categorization of novel objects (for related accounts, see DiCarlo & Cox, 2007; Edelman & Intrator, 1997; Jozwik, Kriegeskorte, & Mur, 2016). Thus, because the feature space has been acquired before the actual experiment, observers should be able to make correct decisions from the first experimental trial on.

A number of previous studies explained classification and learning in terms of generative processes (Feldman, 1992; Feldman, 1997; Kemp, Bernstein, & Tenenbaum, 2005; Lake, Salakhutdinov, & Tenenbaum, 2015; in the domain of semantic knowledge: Jern & Kemp, 2013; Kemp & Jern, 2009; Lake, Salakhutdinov, & Tenenbaum, 2013; Xu & Tenenbaum, 2007). This work tended to focus on identifying models underlying the categorization of abstract stimuli, such as (alphabetical) characters or colored strings produced by grammar-like rules. However, natural objects result from a manifold of physical, chemical, and biological processes that can leave richly structured signatures in object shape that may not necessitate deeper inferences about the (unobservable) underlying generative processes. Indeed, we suggest that in many cases, we can classify complex novel objects using discriminative processes, based on a previously acquired shape representation space for mid-

level shape features—without relying on inferences about the generative process.

To test this hypothesis, we subjected objects to various artificial transformations that introduce complex features into the shape. We used these stimuli to test whether observers could classify objects into groups based on generative processes and whether the responses are based on mid-level representations (i.e., perceptual organization of complex local shape features) rather than Euclidean shape similarity. Specifically, in the first experiment, we test the general question of whether participants can classify unfamiliar objects according to the transformations that have been applied to them. In the second experiment, we test the extent to which such classifications are based on statistical object features rather than Euclidean shape similarity (measured via apparent motion). In the third experiment, we test the extent to which observers can voluntarily distinguish between—and base their classifications on—statistical features associated with transformations, rather than Euclidean shape similarity. Together, these experiments allow us to test whether participants can classify novel objects based on the generative processes that produced them and whether they use the texture-like statistical features that are the hallmarks of the transformations to do so. Note that it does not necessarily follow from good performance in our tasks that participants explicitly infer the causal history of the objects (Schmidt & Fleming, 2018). However, good performance would suggest that generative

processes lead to characteristic features that observers can use as the basis of classification judgments.

## Experiment 1 (three-dimensional shapes)

### Materials and methods

#### Participants

Eight students from Giessen University, 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 procedures were approved by the local ethics board and were carried out in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki). Sample size was chosen based on previous studies; to ensure our findings were not skewed by sample size, we report the Bayes factor for all of our statistical tests (Rouder, Speckman, Sun, Morey, & Iverson, 2009).

#### Stimuli

Eight base objects were handcrafted with Blender 2.76 (Stichting Blender Foundation, Amsterdam, the Netherlands), and each was transformed with eight distinct procedures using Blender sculpting tools, yielding a total of 64 objects. Objects were rendered in Blender (for examples, see Figure 2A; all stimuli can be obtained from <https://doi.org/10.5281/zenodo.2540802>). Factors such as angular size, viewpoint, and illumination were roughly constant for all stimuli to reduce noise from processes not relevant for our research question (e.g., shape from shading, mental rotation). Eight objects (all based on the same base object) were used as match stimuli (i.e., the target categories) and the other 56 objects as test stimuli.

#### Procedure

Stimuli were presented on a black background on a Dell U2412M monitor at a resolution of  $1,920 \times 1,200$  pixels, controlled by MATLAB (MathWorks, Natick, MA) using the Psychophysics Toolbox extension (Kleiner, Brainard, & Pelli, 2007).

On each trial, participants were presented simultaneously with one central test stimulus and the eight surrounding match stimuli (Figure 2A). The match stimuli were all created from the same base object and

were constant across trials. Participants were then asked to choose by a mouse click which match appeared to be subjected to the same transformation as the test stimulus (eight-alternative forced choice [8-AFC] task). Stimuli were presented until response. No feedback was given. Each participant responded to all combinations of test and comparison stimuli with four repetitions ( $56 \times 4$  trials). The order of trials and the assignment of match stimuli to the eight locations around the test stimulus were randomized for each participant.

#### Analysis

We report binomial tests plus the Scaled JZS Bayes factor ( $BF_{10}$ ), using a Jeffrey-Zellner-Siow Prior (Cauchy distribution on effect size) with a default scale factor of 0.707 (Rouder et al., 2009).  $BF_{10}$  expresses the probability of the data given  $H_1$  relative to  $H_0$  (i.e.,  $BF_{10} > 1$  is in favor of  $H_1$ ).  $BF_{10} > 3$  can be considered as “some evidence,”  $BF_{10} > 10$  as “strong evidence,” and  $BF_{10} > 30$  as “very strong evidence” for  $H_1$ , whereas  $BF_{10} < 0.33$  can be considered as “some evidence,”  $BF_{10} < 0.1$  as “strong evidence,” and  $BF_{10} < 0.03$  as “very strong evidence” for  $H_0$  (Jeffreys, 1961). No data were excluded. All data are available for download from <https://doi.org/10.5281/zenodo.2540802>.

### Results

For each of the transformations, we found that performance was well above chance level (0.125) with all binomial tests  $p < 0.001$  (with Bonferroni-corrected significance level  $p < 0.006$ ) and  $BF_{10} > 30$  (Figure 2B). Even when testing correct performance against responses to all other transformations combined, binomial tests for seven of eight transformations yielded  $p < 0.001$  and  $BF_{10} > 30$ . Even for transformation class 8, in response to which participants made the most mistakes (see Figure 2B), the observed data were still roughly four times more likely in favor of the correct answer ( $p = 0.010$ ,  $BF_{10} = 3.94$ ).

To test for learning effects, we fitted a linear regression to the learning curve across trials (Figure 2C), yielding an intercept of 0.84 and a slope of 0.00, strongly suggesting no learning. When testing performance pooled across the first times each transformation was shown, we found evidence for accepting  $H_0$  ( $p = 0.600$ ,  $BF_{10} = 0.13$ , i.e., some evidence tending toward strong evidence, based on only eight trials per participant). Together, these findings suggest performance at the mean level from the very first trial on.

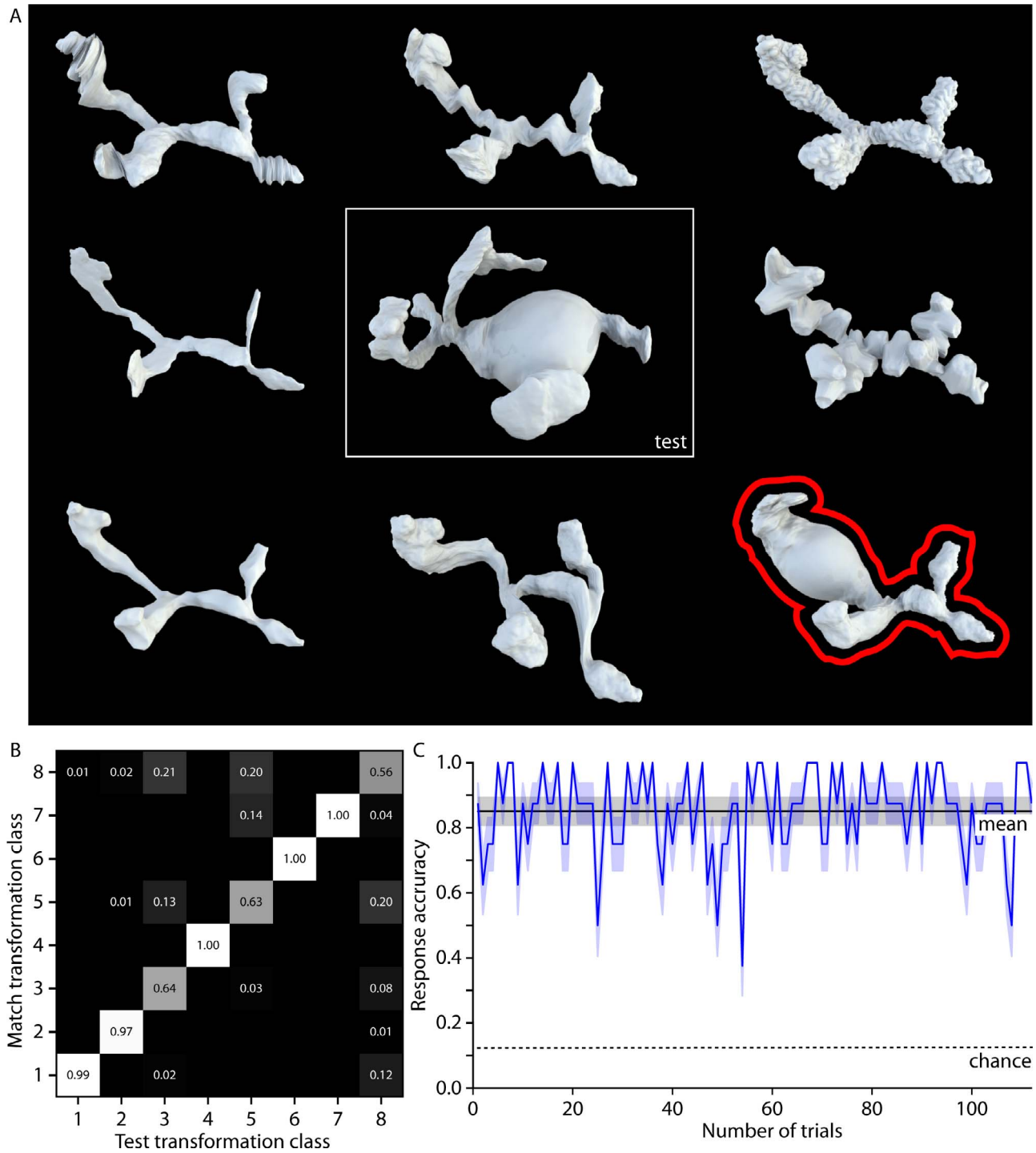


Figure 2. Paradigm and results of Experiment 1. (A) Example trial. Participants were asked to choose from the eight surrounding match stimuli, which was created by the same generative process as the test stimulus in the center (red frame, indicating observers’ correct response). (B) Confusion matrix for the eight transformation classes across all participants; numbers denote proportion correct. (C) Learning curve. The blue line shows the mean performance of participants as a function of experimental trials. Note that performance is already at ceiling from the first trial on. The black solid line shows average performance across participants and trials; the black dotted line shows chance performance. Transparent areas denote standard errors of the mean.

## Experiment 2 (two-dimensional shapes: 2-AFC and apparent motion tasks)

To test more specifically for the role of mid-level texture-like statistical features in the perception of generative processes, we created a new set of two-dimensional objects to pit them against Euclidean shape similarity—a prime source of object classification (e.g., Lowet, Firestone, & Scholl, 2018; Wilder, Feldman, & Singh, 2011). As a control, after asking participants to classify the stimuli, we used apparent motion to measure low-level perceptual shape similarity between stimuli. Specifically, we reasoned that when two objects are presented in series, the degree of apparent motion experienced should increase with the perceptual dissimilarity between the two objects. This allows us to verify that for perceptual processes other than the perception of shape transformations, Euclidean shape similarity—not statistical shape properties—drives similarity judgments.

### Materials and methods

#### Participants

Fifteen students from the Justus-Liebig-University Giessen, Germany, with normal or corrected vision participated in the experiment for financial compensation; all other details were the same as in Experiment 1.

#### Stimuli

We used Adobe Illustrator Effect tools to design four different classes with eight transformed objects each; in addition, for each class, we made stimuli that were similar to the individual transformed objects in pixel-for-pixel terms but designed using a different Effect tool (“Euclidean matches”; 64 stimuli; for examples, see Figure 3A; all stimuli can be obtained from <https://doi.org/10.5281/zenodo.2540802>).

#### Procedure

Stimuli were presented on a white background on a Dell U2412M monitor at a resolution of  $1,920 \times 1,200$  pixels, controlled by MATLAB using the Psychophysics Toolbox extension (Kleiner et al., 2007).

**2-AFC task:** On each trial, participants were presented simultaneously with one central test stimulus, drawn from one of the four transformation classes, and two comparison stimuli (Figure 3C). One comparison stimulus was always the Euclidean match, and the other was always another stimulus of the same transformation class (i.e., different overall shape but

similar statistical features to the test stimulus). Participants were then asked to choose, by a left or right button press, the comparison stimulus that was subjected to the same transformation as the test stimulus (2-AFC task). Specifically, we generated a list of 32 nonsense verbs (e.g., “fumper,” “nattle,” “klompen”), and immediately above the test stimulus was the statement, “This object has been created by a process of Xing,” where “X” was a different nonsense verb on each trial. Underneath the two comparison stimuli was a text reading, “Which of these two objects has also been Xed?” where “X” was the same nonsense word as for the test stimulus. Stimuli were presented until response. Each participant responded to eight blocks with four trials using test stimuli from the four different transformation classes ( $8 \times 4$  trials). The order of blocks, and trials within blocks, as well as the mapping between stimuli and nonsense verbs, was randomized for each participant.

**Apparent motion task:** On each trial, participants were presented with a fixation point at the center of the screen. After 500 ms, two identical test stimuli appeared, one to the left and one to the right side of the fixation point (roughly  $4^\circ$  of visual angle away). After 400 ms, the test stimuli were replaced by two different comparison stimuli (Euclidean match and transformation match), which remained on the screen for 200 ms. Test and comparison stimuli were sorted together as in the 2-AFC task. Participants were asked to choose by a left or right button press whether they perceived more apparent motion for the left or the right stimulus sequence. Each participant responded to two blocks of 32 trials each ( $2 \times 32$  trials). The order of trials within blocks was randomized for each participant.

#### Analysis

We report conventional *t* test plus the Scaled JZS Bayes factor ( $BF_{10}$ ). No data were excluded. All data are available for download from <https://doi.org/10.5281/zenodo.2540802>.

### Results

For the 2-AFC task, we found that responses were clearly biased toward transformation matches (Figure 3B). When testing average responses toward transformation matches versus responses toward Euclidean matches, we found significant and strong evidence for transformation matches in all cases (per transformation: all  $T[14] > 4.83$ ,  $p < 0.001$ , where Bonferroni-corrected significance level requires  $p < 0.012$ , and  $BF_{10} > 30$ ; across transformations:  $T[14] = 9.74$ ,  $p < 0.001$ ,  $BF_{10} > 30$ ; Figure 3B).

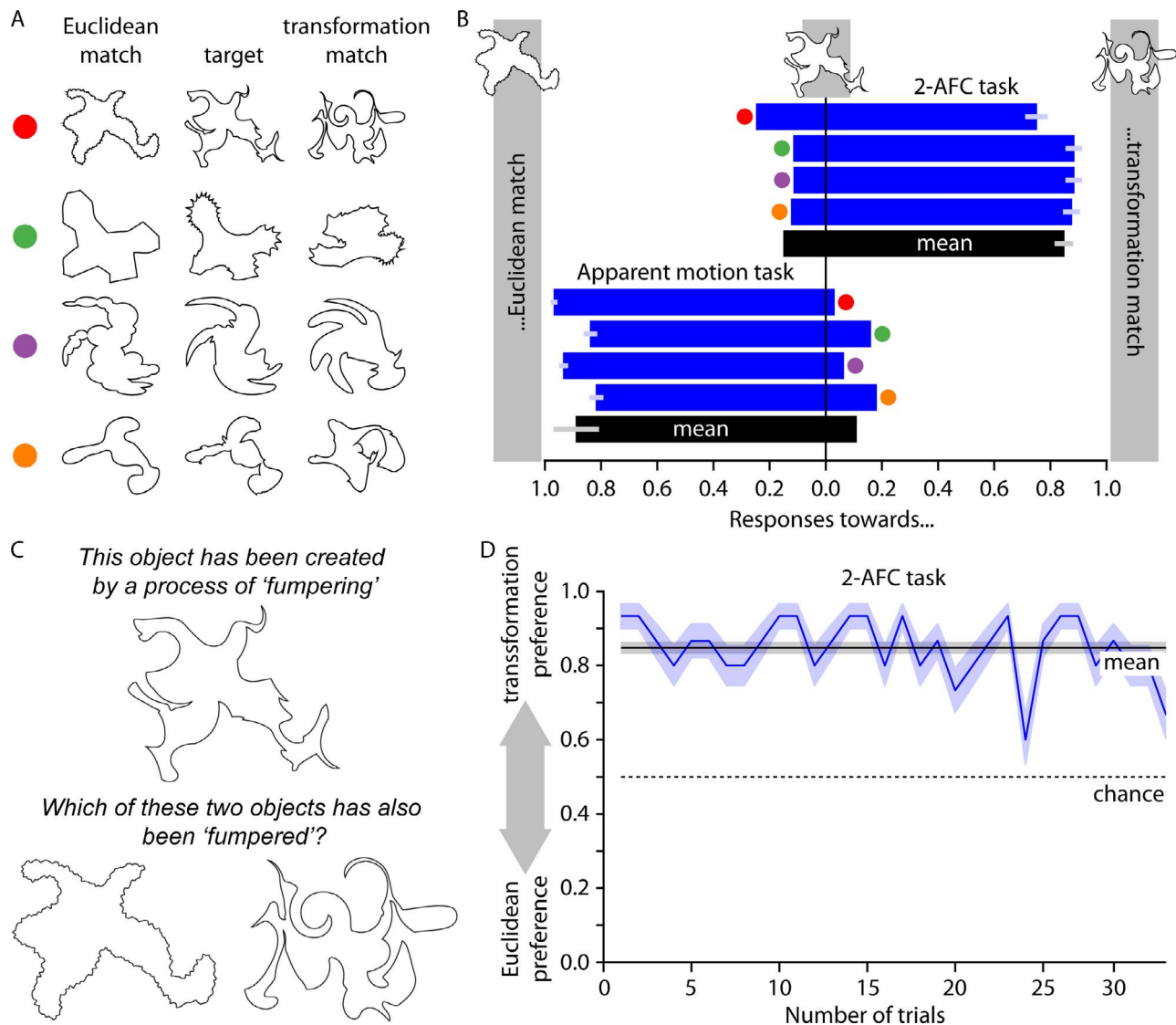


Figure 3. Stimuli and results of Experiment 2. (A) Examples for targets from the four different transformation classes (middle) with Euclidean matches to the left and transformation matches to the right. (B) Overall performance of participants (black bars) and average performance for each of the four transformation classes (blue bars) in the 2-AFC task and the apparent motion task. Performance in both tasks is plotted as a proportion of responses in the direction of the Euclidean match (to the left) versus in the direction of the transformation match (to the right). Bars correspond to the four transformation classes according to the color labeling from (A). Error bars denote standard errors of the mean. (C) Example trial of the 2-AFC task with instructions: The test stimulus is presented in the upper half of the monitor, and the comparison stimuli are presented below (Euclidean match to the left and transformation match to the right). Relative size of stimuli and instructions were different in the actual experiment. (D) Learning curve for the 2-AFC task. The blue line shows the performance of participants as a function of experimental trials. Note that performance is already at ceiling from the first trial on. The black solid line shows the average performance across participants and trials; the black dotted line shows chance performance. Transparent areas denote standard errors of the mean.

For the apparent motion task, the pattern was reversed, with responses clearly biased toward Euclidean matches (per transformation: all  $T[14] > 19.12$ ,  $p < 0.001$ , and  $BF_{10} > 30$ ; across transformations:  $T[14] = 34.21$ ,  $p < 0.001$ ,  $BF_{10} > 30$ ; Figure 3B).

For the 2-AFC task, we also evaluated the learning curve (Figure 3C): A fitted linear regression yielded an

intercept of 0.89 and a slope of 0.00, strongly indicating no learning. When testing performance pooled across the first times each transformation was shown, we found evidence for accepting  $H_0$  ( $T[14] = 0.52$ ,  $p = 0.613$ ,  $BF_{10} = 0.30$ , i.e., some evidence, based on only four trials per participant). This again suggests performance at the mean level from the very first trial on.

## Experiment 3 (two-dimensional shapes: classification task)

In addition to showing that different tasks either favor texture-like statistical features (2-AFC task) or Euclidean shape similarity (apparent motion task), we wanted to test whether participants could access both types of information at will. To do this, we asked two groups of participants to decide whether two stimuli were transformed by the same transformation or whether two stimuli overlapped substantially if they were superimposed on top of one another.

### Materials and methods

#### Participants

Sixteen students participated in the classification task, eight of whom received the transformation instructions, whereas the others received the overlap instructions. All other details were the same as in Experiments 1 and 2.

#### Stimuli

Stimuli were the same as in Experiment 2.

#### Procedure

Stimuli were presented on a white background on a Dell U2412M monitor at a resolution of  $1,920 \times 1,200$  pixels, controlled by MATLAB using the Psychophysics Toolbox extension (Kleiner et al., 2007).

*Classification task: Transformation instructions:* On each trial, participants were presented simultaneously with one of the stimuli of the transformations classes and its corresponding Euclidean match. Participants made a binary judgment about whether the two stimuli were subjected to the same transformation or not. Stimuli were presented until response. Each participant responded to all 480 combinations (120 combinations per transformation class) with two repetitions ( $480 \times 2$  trials). The order of trials was randomized for each participant.

*Classification task: Overlap instructions:* The task was the same as above, except that participants were asked whether or not the two stimuli would overlap substantially if they were superimposed on top of one another.

#### Analysis

We report the  $R^2$  from correlations between the predictions (Figure 3A) and our data (Figure 3B) plus the corresponding Scaled JZS Bayes factor ( $BF_{10}$ ) with

a default scale factor of 0.707 (Wetzels & Wagenmakers, 2012). No data were excluded. All data are available for download from <https://doi.org/10.5281/zenodo.2540802>.

### Results

For the transformation instructions, responses were clearly biased toward the transformation relevant prediction, and for the overlap instructions, responses were clearly biased toward the Euclidean match relevant prediction (Figure 4). We correlated the similarity matrices—containing all 16 transformation stimuli and their Euclidean matches (yielding 136 comparisons) for each of the four transformation classes. For the transformation instructions, we found strong correlations with the transformation relevant prediction (all  $r[135] > 0.96$ ,  $R^2 > 0.93$ ,  $p < 0.001$ , and  $BF_{10} > 30$ ; overall:  $r[543] > 0.98$ ,  $R^2 > 0.96$ ,  $p < 0.001$ , and  $BF_{10} > 30$ ) and weak correlations with the Euclidean match relevant prediction (all  $r[135] < 0.24$ ,  $R^2 < 0.06$ ,  $p < 0.006$ , and  $BF_{10} > 0.84$ ; overall:  $r[543] < 0.21$ ,  $R^2 < 0.05$ ,  $p < 0.001$ , and  $BF_{10} > 30$ ; red dots in Figure 3C). For the overlap instructions, we found a strong correlation with the Euclidean match relevant prediction (all  $r[135] > 0.98$ ,  $R^2 > 0.94$ ,  $p < 0.001$ , and  $BF_{10} > 30$ ; overall:  $r[543] > 0.97$ ,  $R^2 > 0.96$ ,  $p < 0.001$ , and  $BF_{10} > 30$ ) and weak correlations with the transformation relevant prediction (all  $r[135] < 0.28$ ,  $R^2 < 0.08$ ,  $p < 0.031$ , and  $BF_{10} > 0.69$ ; overall:  $r[543] < 0.23$ ,  $R^2 < 0.06$ ,  $p < 0.001$ , and  $BF_{10} > 30$ ; green dots in Figure 3C).

### Discussion

Whether two objects belong to the same or different categories is intimately related to the generative processes that brought them into being. We hypothesized that in many cases, shape-altering transformations leave highly distinctive statistical signatures in object shapes, which the visual system could exploit for categorizing novel objects. Our findings demonstrate that observers are indeed excellent at identifying objects that share common generative processes. In Experiment 1, participants performed far above chance at classifying unfamiliar three-dimensional objects based on the transformations that had been applied to them. They could do this from the very first trials on, indicating that they solve this task with complex novel objects by relying on a previously established shape feature space. This is in line with previous research on simple objects (e.g., Smith, Jones, Landau, Gershkoff-Stowe, & Samuelson, 2002), characters (Lake et al.,

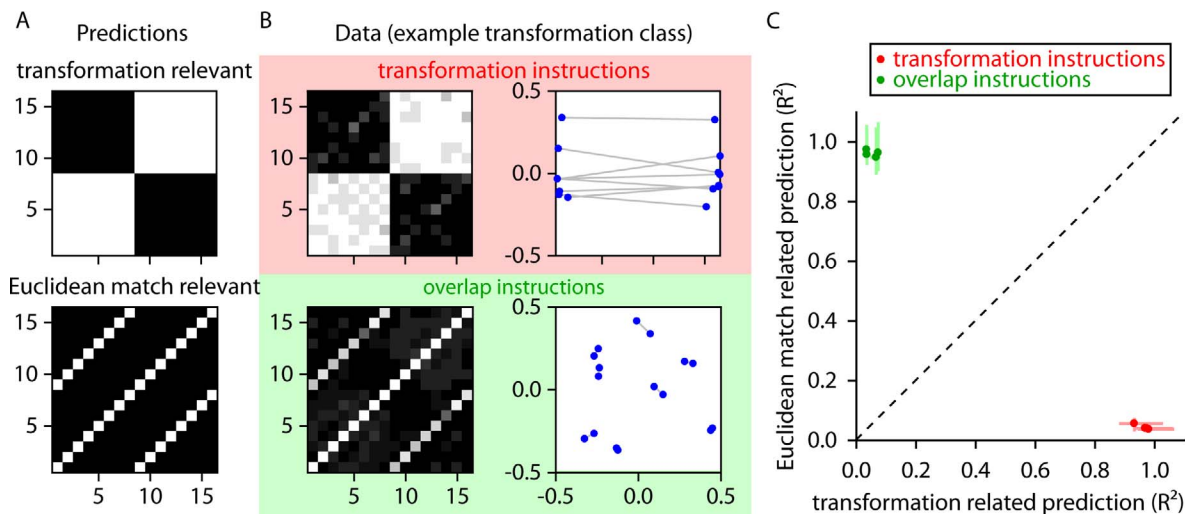


Figure 4. Results of Experiment 3. (A) Predictions for similarity ratings when only transformation is relevant (upper row) versus when only Euclidean match is relevant (lower row). Similarity matrices show similarity for all 16 transformation stimuli of one transformation class and their Euclidean matches. (B) Exemplary data for one of the four transformation classes as obtained with the transformation instructions (upper row) and overlap instructions (lower row). The left column (similarity matrices) shows average classification responses; the right column shows all stimuli plotted in multi-dimensional scaling space (gray lines connect each transformation stimulus with its Euclidean match). (C) Results for all four transformation classes were correlated with the predictions for only transformation relevant (x-axis) and only Euclidean match relevant (y-axis), and resulting  $R^2$  values were plotted in the space spanned by these two axes. Results from the transformation instructions (red) and the overlap instructions (green) are plotted separately for the four transformation classes (transparent error bars denote 95% confidence intervals).

2013) and words (Jern & Kemp, 2013; Kemp & Jern, 2009; Xu & Tenenbaum, 2007), demonstrating a general capability of humans to make categorizations based on single exemplars. It does not follow from our results that participants explicitly infer the generative processes that shaped the objects. Instead, we suggest that the participants' ability to classify objects is a result of their ability to discriminate stimuli using the statistical shape features that are associated with different generative processes.

Experiment 2 demonstrated that sensitivity to the statistical signatures of transformations is distinct from Euclidean shape similarity. Identical stimuli yielded opposite response patterns across tasks: Although apparent motion was driven by the distances between local contour features across shapes, inferences about causality (2-AFC task) were based on texture-like statistical shape features. Experiment 3 revealed that participants can voluntarily separate and respond to these different aspects of the shape. When asked about generative processes, they grouped objects together based on statistical features; when asked about the extent of overlap, they grouped objects based on raw pixel similarities.

Together, these findings demonstrate that shape provides rich cues for categorizations via the statistical traces left in objects by processes in their past. The key insight is that the visual system represents object shape using a very large number of distinct perceptual features,

which define a high-dimensional “shape space” (e.g., DiCarlo & Cox, 2007; Leeds, Pyles, & Tarr, 2014). We speculate that objects created by similar generative processes tend to lie much closer together in the shape space than objects created by different generative processes (Figure 5). Thus, the visual system could use a simple heuristic based on distances between items to determine whether two unfamiliar objects belong to the same class. The distance threshold for the same/different decision could be learnt from the distributions of items within familiar categories (i.e., a prior on the typical distances between items within categories).

In a similar vein, Jern and Kemp (2013) demonstrated, with simple arrow stimuli varying in length, color and saturation, that sampling from the observed probability distributions of object features over exemplars and categories enables the creation of new exemplars for these categories. Salakhutdinov, Tenenbaum, and Torralba (2012) proposed a hierarchical probabilistic model that transfers acquired knowledge from previously learned categories to a novel category. The model can to some extent infer novel classes from single category examples: priors derived from previous knowledge about “super categories” are used to estimate prototypes and features of the new class. We speculate that for richly structured shapes—like the ones we used—simple discriminative models in a high-dimensional shape space may be sufficient to account for putative “one-shot” learning.



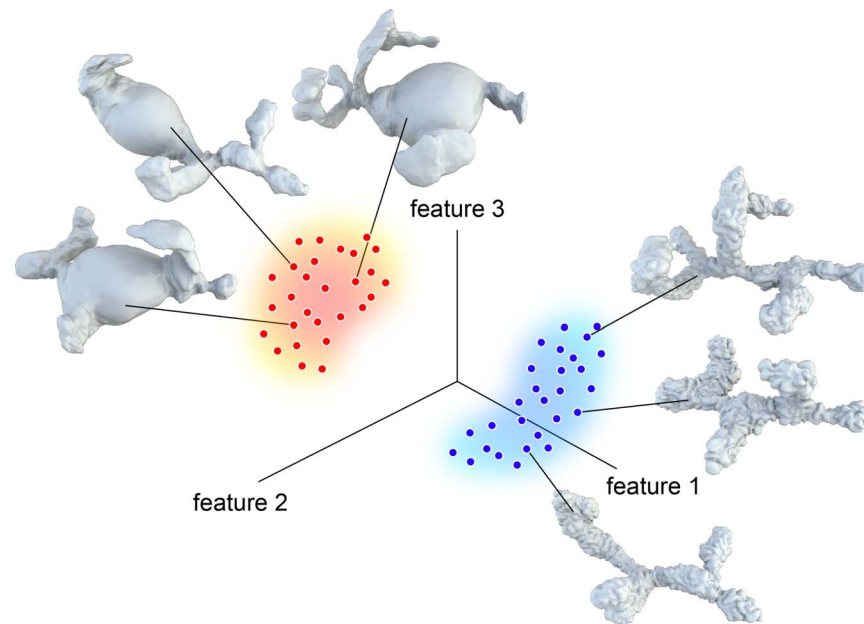


Figure 5. Two classes of objects with distinct generative histories separated from one another in a hypothetical perceptual “shape space” (depicted here with only three dimensions). Given a large enough number of appropriate feature dimensions, objects within a class will tend to be closer to one another than to members of any other class. Thus, given a pair of novel objects, the distance between them could be used to determine whether they belong to the same class.

Crucial to this approach, however, is the choice of features that define the perceptual shape space: Similar generative processes should lead to similar locations in shape space. However, similar generative processes do not necessarily create shapes that are similar to one another on a point-by-point basis. For example, the objects in Experiment 1 have multiple limbs with a variety of positions and curvatures, which do not line up with one another across different instances within each class. Nevertheless, the “twisting” transformation leads to distinctive spiral-shaped features on the limbs, which the visual system could use for identifying objects that have been subjected to the same transformation. We argue that at least some of the features that define shape space are the result of “mid-level” perceptual organization computations that describe relationships between multiple distant locations of the object (cf. Jozwik et al., 2016; van Assen, Barla, & Fleming, 2018). Thus, texture-like representations of statistical shape features likely play an important role in inferences related to causal history.

Previous studies can help us to figure out the details of this process. Op de Beeck, Torfs, and Wagemans (2008) measured perceived similarity for a set of objects in which they independently varied overall shape envelope (e.g., “square” and “vertical” objects) and local shape features (e.g., “spiky” and “smooth” objects). They found that participants judge similarity based on both factors; that is, they perceived objects as more similar if they shared either overall shape, local shape features, or both. Neural activation as measured

by functional magnetic resonance imaging was correlated with these similarity ratings. Specifically, similarity in overall shape was linked to activation in retinotopic areas lower in the visual hierarchy (V1, V2, V3, V4v), whereas similarity in local shape features was linked to activation in nonretinotopic area lateral occipital cortex (LOC) higher in the visual hierarchy. This is also supported by work from Kubilius, Bracci, and Op de Beeck (2016) showing that deep computational models predict object similarities in human perception, with lower levels of the networks mainly representing similarities in overall shape and higher levels representing similarities in local shape features (Elder, 2018). These findings suggest that classification based on causal history is mediated by the analysis of local shape features in higher-level shape representations.

This is in line with a layered view of shape perception in which objects consist of multiple shape properties at varying degrees of abstraction (Green, 2015), for example, parallel representations of overall shape and local shape features, which are retrieved depending on the task at hand (as in our overlap vs. transformation instructions in Experiment 3) and potentially can be computed very fast (Epshtein, Lifshitz, & Ullman, 2008). This view provides a means to unify transformational and generative approaches to object categorization. Transformational approaches (e.g., Bedford & Mansson, 2010; Hahn, Chater, & Richardson, 2003; Hahn, Close, & Graf, 2009; Imai, 1977) argue that similarity judgments depend on transformational distance between objects (i.e., overall shape similarity).

Generative approaches (e.g., Kemp et al., 2005) emphasize inferences about common generative processes (i.e., similarity in statistical shape features). Our results show that there are multiple levels of shape representation that the visual system draws on depending on the task and available information.

Evidence from child development research suggests that the ability to access multiple levels of shape representation is available relatively early in life. When 4-year-old children were presented with one standard object of a particular overall shape and texture, they grouped a novel object based on shape rather than texture; however, when presented with two standard objects of different shapes but similar textures, they grouped a novel object based on texture rather than shape (Graham, Namy, Gentner, & Meagher, 2010). This selection of features is also affected by semantic cues: for example, 3-year-old children grouped objects based on superordinate characteristics (e.g., animals vs. food) when objects were labeled with nouns (e.g., “momos”), but they grouped objects based on subordinate characteristics (e.g., relying on color or texture; red grapes vs. green grapes) when objects were labeled with adjectives (e.g., “mom-ish” ones; Waxman, 1990).

To conclude, we found that observers can categorize complex novel objects by perceiving the signatures of generative processes, even from the very first experimental trials on. These findings also have implications for the perception of causal history and shape similarity judgments. Specifically, the underlying shape representations likely aid inferences about the causal history of objects (e.g., Leyton, 1989; Pinna, 2010; Pinna & Deiana, 2015; Spröte & Fleming, 2013), and objects that share features produced by the same generative process should appear more similar compared with other shapes (e.g., Ons & Wagemans, 2012; Op de Beeck et al., 2008). Finally, these visual processes could potentially facilitate other shape and material perception tasks, including (a) identifying the physical properties of objects (Paulun, Schmidt, van Assen, & Fleming, 2017; Schmidt, Paulun, van Assen, & Fleming, 2017; van Assen et al., 2018), (b) making predictions about what other members of the same category might look like (i.e., mental imagery of “plausible variants”), (c) motor affordances, and (d) predicting future states of moving and interacting objects.

*Keywords:* shape, perceptual organization, categorization, generalization, causal history

## Acknowledgments

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research

Foundation; project number 222641018–SFB/TRR 135 TP C1) and ERC Consolidator Award “SHAPE” (ERC-CoG-2015-682859).

R.W.F. developed study concept and design. Both authors collected data, performed analyses and interpretation, and wrote the manuscript. Both authors approved the final version for submission.

Commercial relationships: none.

Corresponding author: Filipp Schmidt.

Email: filipp.schmidt@psychol.uni-giessen.de.

Address: Justus-Liebig-University Giessen, General Psychology, Gießen, Germany.

## References

- Arnheim, R. (1974). *Art and visual perception: A psychology of the creative eye*. Berkeley, CA: University of California Press.
- Bedford, F. L., & Mansson, B. E. (2010). Object identity, apparent motion, transformation geometry. *Current Research in Psychology, 1*, 35–52, <https://doi.org/10.3844/crpsp.2010.35.52>.
- Chen, Y.-C., & Scholl, B. J. (2016). The perception of history: Seeing causal history in static shapes induces illusory motion perception. *Psychological Science, 27*, 923–930, <https://doi.org/10.1177/0956797616628525>.
- DiCarlo, J. J., & Cox, D. D. (2007). Untangling invariant object recognition. *Trends in Cognitive Sciences, 11*, 333–341, <https://doi.org/10.1016/j.tics.2007.06.010>.
- Edelman, S., & Intrator, N. (1997). Learning as formation of low-dimensional representation spaces. In G. W. Cottrell (Ed.), *Proceedings of the 18th Annual Conference of the Cognitive Science Society* (pp. 199–204). London: Psychology Press.
- Elder, J. H. (2018). Shape from contour: Computation and representation. *Annual Review of Vision Science, 4*, 423–450, <https://doi.org/10.1146/annurev-vision-091517-034110>.
- Epshtein, B., Lifshitz, I., & Ullman, S. (2008). Image interpretation by a single bottom-up top-down cycle. *Proceedings of the National Academy of Sciences, USA, 105*, 14298–14303, <https://doi.org/10.1073/pnas.0800968105>.
- Feldman, J. (1992). Constructing perceptual categories. In: *Computer vision and pattern recognition: Conference: Papers* (pp. 244–250). Washington, DC: IEEE Computer Society Press, <https://doi.org/10.1109/CVPR.1992.223268>.

- Feldman, J. (1997). The structure of perceptual categories. *Journal of Mathematical Psychology*, *41*, 145–170, <https://doi.org/10.1006/jmps.1997.1154>.
- Graham, S. A., Namy, L. L., Gentner, D., & Meagher, K. (2010). The role of comparison in preschoolers' novel object categorization. *Journal of Experimental Child Psychology*, *107*, 280–290, <https://doi.org/10.1016/j.jecp.2010.04.017>.
- Green, E. J. (2015). A layered view of shape perception. *British Journal for the Philosophy of Science*, *68*, 355–387, axv042, <https://doi.org/10.1093/bjps/axv042>.
- Hahn, U., Chater, N., & Richardson, L. B. (2003). Similarity as transformation. *Cognition*, *87*, 1–32, [https://doi.org/10.1016/S0010-0277\(02\)00184-1](https://doi.org/10.1016/S0010-0277(02)00184-1).
- Hahn, U., Close, J., & Graf, M. (2009). Transformation direction influences shape-similarity judgments. *Psychological Science*, *20*, 447–454, <https://doi.org/10.1111/j.1467-9280.2009.02310.x>.
- Imai, S. (1977). Pattern similarity and cognitive transformations. *Acta Psychologica*, *41*, 433–447, [https://doi.org/10.1016/0001-6918\(77\)90002-6](https://doi.org/10.1016/0001-6918(77)90002-6).
- Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford, UK: Oxford.
- Jern, A., & Kemp, C. (2013). A probabilistic account of exemplar and category generation. *Cognitive Psychology*, *66*, 85–125, <https://doi.org/10.1016/j.cogpsych.2012.09.003>.
- 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, <https://doi.org/10.1016/j.neuropsychologia.2015.10.023>.
- Kemp, C., Bernstein, A., & Tenenbaum, J. B. (2005). A generative theory of similarity. In B. G. Bara, L. Barsalou, & M. Bucciarelli (Eds.), *Proceedings of the 27th Annual Meeting of the Cognitive Science Society* (pp. 1132–1137). Austin, TX: Cognitive Science Society.
- Kemp, C., & Jern, A. (2009). Abstraction and relational learning. In Y. Bengio, D. Schuurmans, J. D. Lafferty, C. K. I. Williams, & A. Culotta (Eds.), *Advances in neural information processing systems 22* (pp. 934–942). Red Hook, NY: Curran Associates.
- Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3? *Perception*, *36*, 1, <https://doi.org/10.1068/v070821>.
- Kubilius, J., Bracci, S., & Op de Beeck, H. P. (2016). Deep neural networks as a computational model for human shape sensitivity. *PLoS Computational Biology*, *12*, e1004896, <https://doi.org/10.1371/journal.pcbi.1004896>.
- Lake, B. M., Salakhutdinov, R., & Tenenbaum, J. B. (2013). One-shot learning by inverting a compositional causal process. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, & K. Q. Weinberger (Eds.), *Advances in neural information processing systems 26* (pp. 2526–2534). Red Hook, NY: Curran Associates. Retrieved from <http://papers.nips.cc/paper/5128-one-shot-learning-by-inverting-a-compositional-causal-process.pdf>
- Lake, B. M., Salakhutdinov, R., & Tenenbaum, J. B. (2015, December 11). Human-level concept learning through probabilistic program induction. *Science*, *350*, 1332–1338, <https://doi.org/10.1126/science.aab3050>.
- Leeds, D. D., Pyles, J. A., & Tarr, M. J. (2014). Exploration of complex visual feature spaces for object perception. *Frontiers in Computational Neuroscience*, *8*, 106, <https://doi.org/10.3389/fncom.2014.00106>.
- Leyton, M. (1989). Inferring causal history from shape. *Cognitive Science*, *13*, 357–387, [https://doi.org/10.1207/s15516709cog1303\\_2](https://doi.org/10.1207/s15516709cog1303_2).
- Lowet, A. S., Firestone, C., & Scholl, B. J. (2018). Seeing structure: Shape skeletons modulate perceived similarity. *Attention, Perception & Psychophysics*, *80*, 1278–1289, <https://doi.org/10.3758/s13414-017-1457-8>.
- Ons, B., & Wagemans, J. (2012). Generalization of visual shapes by flexible and simple rules. *Seeing and Perceiving*, *25*, 237–261, <https://doi.org/10.1163/187847511X571519>.
- Op de Beeck, H. P., Torfs, K., & Wagemans, J. (2008). Perceived shape similarity among unfamiliar objects and the organization of the human object vision pathway. *Journal of Neuroscience*, *28*, 10111–10123, <https://doi.org/10.1523/JNEUROSCI.2511-08.2008>.
- Paulun, V. C., Schmidt, F., van Assen, J. J. R., & Fleming, R. W. (2017). Shape, motion, and optical cues to stiffness of elastic objects. *Journal of Vision*, *17*(1):20, 1–22, <https://doi.org/10.1167/17.1.20>. [PubMed] [Article]
- Pinna, B. (2010). New Gestalt principles of perceptual organization: An extension from grouping to shape and meaning. *Gestalt Theory*, *32*, 11–78.
- Pinna, B., & Deiana, K. (2015). Material properties from contours: New insights on object perception. *Vision Research*, *115*(Pt B), 280–301, <https://doi.org/10.1016/j.visres.2015.03.014>.
- Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian *t* tests for accepting

- and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*, 225–237, <https://doi.org/10.3758/PBR.16.2.225>.
- Salakhutdinov, R., Tenenbaum, J. B., & Torralba, A. (2012). One-shot learning with a hierarchical nonparametric Bayesian model. *JMLR: Workshop and Conference Proceedings*, *27*, 195–207.
- Schmidt, F., & Fleming, R. W. (2016). Visual perception of complex shape-transforming processes. *Cognitive Psychology*, *90*, 48–70, <https://doi.org/10.1016/j.cogpsych.2016.08.002>.
- Schmidt, F., & Fleming, R. W. (2018). Identifying shape transformations from photographs of real objects. *PLoS One*, *13*, e0202115, <https://doi.org/10.1371/journal.pone.0202115>.
- Schmidt, F., Paulun, V. C., van Assen, J. J. R., & Fleming, R. W. (2017). Inferring the stiffness of unfamiliar objects from optical, shape, and motion cues. *Journal of Vision*, *17*(3):18, 1–17, <https://doi.org/10.1167/17.3.18>. [PubMed] [Article]
- Smith, L. B., Jones, S. S., Landau, B., Gershkoff-Stowe, L., & Samuelson, L. (2002). Object name learning provides on-the-job training for attention. *Psychological Science*, *13*, 13–19, <https://doi.org/10.1111/1467-9280.00403>.
- Spröte, P., & Fleming, R. W. (2013). Concavities, negative parts, and the perception that shapes are complete. *Journal of Vision*, *13*(14):3, 1–23, <https://doi.org/10.1167/13.14.3>. [PubMed] [Article]
- Spröte, P., Schmidt, F., & Fleming, R. W. (2016). Visual perception of shape altered by inferred causal history. *Scientific Reports*, *6*, 36245, <https://doi.org/10.1038/srep36245>.
- Van Assen, J. J. R., Barla, P., & Fleming, R. W. (2018). Visual features in the perception of liquids. *Current Biology*, *28*, 452–458. e4, <https://doi.org/10.1016/j.cub.2017.12.037>.
- Waxman, S. R. (1990). Linguistic biases and the establishment of conceptual hierarchies: Evidence from preschool children. *Cognitive Development*, *5*, 123–150, [https://doi.org/10.1016/0885-2014\(90\)90023-M](https://doi.org/10.1016/0885-2014(90)90023-M).
- Wetzels, R., & Wagenmakers, E.-J. (2012). A default Bayesian hypothesis test for correlations and partial correlations. *Psychonomic Bulletin & Review*, *19*, 1057–1064, <https://doi.org/10.3758/s13423-012-0295-x>.
- Wilder, J., Feldman, J., & Singh, M. (2011). Superordinate shape classification using natural shape statistics. *Cognition*, *119*, 325–340, <https://doi.org/10.1016/j.cognition.2011.01.009>.
- Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review*, *114*, 245–272, <https://doi.org/10.1037/0033-295X.114.2.245>.