Visual Perception of Shape-Transforming Processes: 'Shape Scission'

Filipp Schmidt[1], Flip Phillips[3], & Roland W. Fleming[1,2]

[1] Justus Liebig University Giessen, Germany
[2] Center for Mind, Brain and Behavior, Giessen, Germany
[3] Skidmore College, NY, USA

This is a post-peer-review, pre-copyedit version of an article published in ***Cognition.*** *The final authenticated version is available online at* https://doi.org/10.1016/j.cognition.2019.04.006

Abstract

Shape-deforming processes (e.g., squashing, bending, twisting) can radically alter objects' shapes. After such a transformation, some features are due to the object's original form, while others are due to the transformation, yet it is challenging to separate the two. We tested whether observers can distinguish the causal origin of different features, teasing apart the characteristics of the original shape from those imposed by transformations, a process we call 'shape scission'. Using computer graphics, we created 8 unfamiliar objects and subjected each to 8 transformations (e.g., "twisted", "inflated", "melted"). One group of participants named transformations consistently. A second group arranged cards depicting the objects into classes according to either (i) the original shape or (ii) the type of transformation. They could do this almost perfectly, suggesting that they readily distinguish the causal origin of shape features. Another group used a digital painting interface to indicate which locations on the objects appeared transformed, with responses suggesting they can localise features caused by transformations. Finally, we parametrically varied the magnitude of the transformations, and asked another group to rate the degree of transformation. Ratings correlated strongly with transformation magnitude with a tendency to overestimate small magnitudes. Responses were predicted by both the magnitude and area affected by the transformation. Together, the findings suggest that observers can scission object shapes into original shape and transformation features and access the resulting representational layers at will.

Keywords: vision; objects; recognition; categorization; perceptual organization; Gestalt

1 Introduction

Every object in our environment is the result of generative processes such as manufacture, biological growth, or self-organization in response to physical forces (e.g., gravity or temperature). These processes shape objects. For example, dough is *kneaded* into a croissant, a shoot *grows* into a tree, a tent *droops* under gravity, and a scoop of ice cream *melts* into a puddle. The transformation of object shape poses a challenge to our visual and cognitive systems, which have to solve two complementary and linked inferences: recognizing objects across transformations, and recognizing transformations across objects. While there is a huge literature on object recognition (Biederman, 1987; DiCarlo, Zoccolan, & Rust, 2012; Logothetis & Sheinberg, 1996; Pasupathy, El-Shamayleh, & Popovkina, 2018; Riesenhuber & Poggio, 2002), far less attention has been paid to the recognition of transformations (Arnheim, 1974; Leyton, 1989; Mark & Todd, 1985; Ons & Wagemans, 2012; Pinna, 2010; Pinna & Deiana, 2015; Pittenger & Todd, 1983; Schmidt & Fleming, 2018; Spröte & Fleming, 2013, 2016). Here we investigate the perception of object transformations, focusing on the extent to which participants can determine the causal origin of different shape features (Fig. 1).

Specifically, when presented with a single, static image of an unfamiliar object, there is a fundamental ambiguity about the origin of its observable features. Some features may 'belong' to the object, while others may be imposed or added by a transformation that has altered it from its original form. Here, we investigate the extent to which participants can tease apart these different causal origins, distinguishing shape features that 'belong' to the object from those that 'belong' to a shape-altering transformation applied to the object. We call this process 'shape scission' (Schmidt & Fleming, 2018; Spröte & Fleming, 2016).

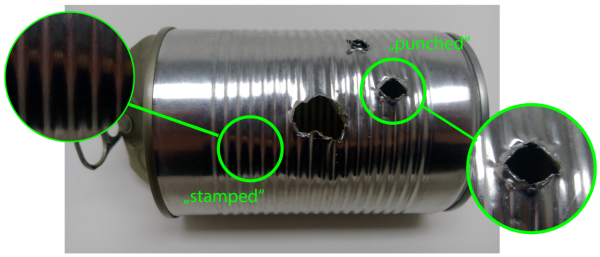

Figure 1. Example of a transformed object in which different shape features are produced by different causal origins–such as the stamped ridges or the punched holes.

1.1 Related work

When presented with a pair of objects that are related to one another by some kind of transformation, how well can we identify features that the objects have in common? A number of studies have investigated the extent to which observers can identify corresponding features or points across transformations. The findings suggest that observers are remarkably good at determining correspondences across a wide range of 2D and 3D transformations, including complex non-rigid deformations (Cornelis, van Doorn, & Wagemans, 2009; Koenderink, van Doorn, Kappers, & Todd, 1997; Phillips, Todd, Koenderink, & Kappers, 1997; Phillips, Todd, Koenderink, & Kappers, 2003; Schmidt & Fleming, 2016; Schmidt, Spröte, & Fleming, 2016; Ward, Isik, & Chun, 2018).

A more challenging case is when only a single object is presented, and observers are asked to infer object's *causal history* (Leyton, 1989) without any reference figure to compare against (Arnheim, 1974; Mark & Todd, 1985; Pinna, 2010; Pinna & Deiana, 2015; Pittenger & Todd, 1983; Schmidt & Fleming, 2018; Spröte & Fleming, 2013, 2016). Despite the fact that this is an under-constrained task, under appropriate circumstances, we can make certain inferences about the processes that have shaped an object in its past. For example, the croissant in Figure 2B, appears to have been bent. Pinna (2010) showed that observers consistently assigned specific meanings to line drawings of simple shapes subjected to different transformations (e.g., "gnawed" or "deliquescing" squares). Our own previous findings suggest that observers can also (i) match the amount of bending between

different objects and (ii) discount the effects of bending for single objects (Spröte & Fleming, 2016). There is also evidence that the inference of causal history can affect the judgment of other shape characteristics (such as symmetry axis and front and back; Spröte, Schmidt, & Fleming, 2016) and the perception of illusory motion (Chen & Scholl, 2016).

Notably, Chen and Scholl (2016) also reported a distinct effect of inferred causal history on visual processing. Specifically, they asked observers to report whether a change of a square to a truncated square (with a piece missing) was sudden or gradual. They found that sudden changes were to some extent perceived as gradual–but only when the resulting truncated square was consistent with a causal intrusion or extrusion event (e.g., when a disc "pushed" into the square was followed by an indentation in the shape of an arch entrance rather than in the shape of a disc). Therefore, the illusory motion demonstrated that observers saw events differently depending on inferred causal history. This is a strong example of causal history impacting perceptual processes, although more generally both perception and cognition—including, for example mental imagery, or even explicit reasoning—may be involved in inferences about causal history.

1.2 Shape Scission

Here, we suggest that one useful way to pose inferences about transformations and causal history is in terms of a process we call *shape scission*. The term 'scission' is typically associated with the perception of transparency, in which the visual system must separate a single retinal intensity or colour value into two distinct surface layers, visible along a single line of sight (Anderson, 1997; Anderson, 2003; Koffka, 1935/1965; Masin, 1999; Metelli, 1970). However, it also used to describe blind source separation problems more generally, such as the separation of retinal luminance values into reflectance and illumination to achieve lightness constancy (e.g., Anderson & Winawer, 2005; Kanizsa, 1955; Metelli, 1985; Singh & Anderson, 2002; cf. *intrinsic image analysis*, Barrow & Tenenbaum, 1978) (Fig. 2A). In scission models of perception, the sources are represented as separate layers capturing the contribution of each source, as in intrinsic image analysis. By analogy, 'shape scission' describes the separation of the observed shape into separate layers representing (i) the original shape, and (ii) the transformation that produced the observed shape from the original shape (Fig. 2B). In other words, the process of 'shape scission' distinguishes between intrinsic contributions (e.g., typical shape, material properties) and extrinsic contributions (e.g., orientation, lighting, causal history) to the observed shape. In the limit, if the visual system were able to do this perfectly, we would end up with two causal layers representing the intrinsic and extrinsic factors, allowing us to fully recover both the original shape and the causal history or transformation. Of course, in practice, we rarely have a vivid subjective experience of the original shape or exactly what has been done to it (e.g., we cannot perceptually invert the complex pattern of folds and creases in a crumpled ball of paper). However, we suggest that under many circumstances we can at least partially separate features of objects based on their causal origin. Note that in contrast to other scission processes, 'shape scission' is not necessarily an unconscious perceptual inference but may involve cognitive as well as visual components.

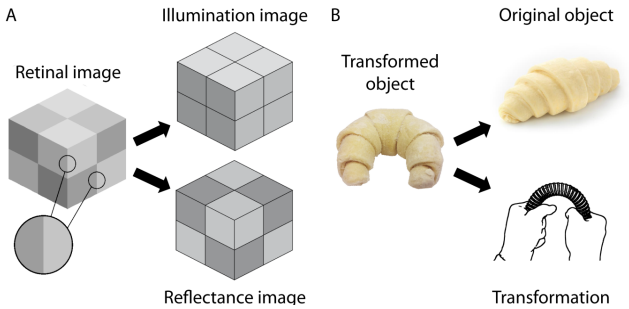

Figure 2. (A) Illustration of the scission problem in lightness perception (inspired by Adelson, 2000). (B) Illustration of the scission problem in shape perception.

1.3 Current Study

Probing the exact nature of shape representations in the brain is challenging, but we reasoned that if the visual system can at least partially scission shape, it should enable participants to perform a number of tasks that otherwise, due to their ill-posed nature, would be difficult or impossible to

perform. Specifically, we sought to test whether observers can: (1) identify transformations that have been applied to objects, (2) separate features by their causal origin, (3) localize features associated with shape-altering transformations, and (4) estimate the magnitude of transformations applied to objects. To this end, we produced a set of handcrafted, novel 3D objects and tested these different aspects of 'shape scission' by asking observers to access information in different causal layers. Specifically, we asked them to identify transformations and assign meaning to them, classify objects according to their original shape or according to transformations, indicate transformed regions on the object via a painting interface, and estimate the overall transformation magnitude. This set of experiments goes beyond previous studies in several key ways: (1) we investigate a broader range of transformations with parametric variations of the magnitude of transformation. (2) We use carefully controlled stimuli with factorial combinations of original shape and transformation with other factors held constant. (3) We test multiple tasks on the same stimulus set, providing convergent evidence for 'shape scission'. (4) Most importantly, the use of transformations with localizable effects enables comparisons between human judgments and computational measures applied directly to the underlying 3D mesh representations, paving the way for the development of image-computable models for predicting human judgments.

2 Experiment 1 (Free Naming)

The first experiment was conducted to validate our stimuli. As the stimuli were handcrafted by us, we wanted to make sure that they actually resemble real-world transformations. To test this, we employed a free naming task in which participants were asked to provide names describing the transformations that had been applied to the objects.

2.1 Materials and Methods

2.1.1 Participants. 16 students from 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 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).

2.1.2 Stimuli. For Experiment 1, we created images of 64 transformed objects, defined by eight different transformations applied to eight novel 3D base objects. The details are described in the following. All stimuli can be obtained from https://doi.org/10.5281/zenodo.2609016.

2.1.2.1 Base objects and transformations. We used Blender 2.76 (Stichting Blender Foundation, Amsterdam, NL), an open-source 3D computer graphics application, to create eight irregular base objects (Fig. 3A). Then, we created for each of these base objects eight different versions with localized transformations, using Blender Sculpt Tools (built-in tools plus "Dried Ground Brush" by DennisH2010, "Custom brushes" by vAonom, and "OrbBrushesPack" by stkopp) (Fig. 3B). Note that in Experiment 1-3, participants were never shown the base objects.

2.1.2.2 Renderings. The render engine used to generate the final images was Maxwell (V. 3.0.1.3; NextLimit Technologies, Madrid, Spain). All base objects and transformed objects were rendered with a blue plastic material, that in a previous study was judged to be of intermediate softness (Schmidt, Paulun, van Assen, & Fleming, 2017). The ground plane was rendered with a textured grey surface. The images were rendered at a resolution of 800 × 600 pixels, a sampling level of 18, and scenarios were lighted by a studio-like environment map. A selection of stimuli rendered for Experiments 1-3 is depicted in Figure 3.

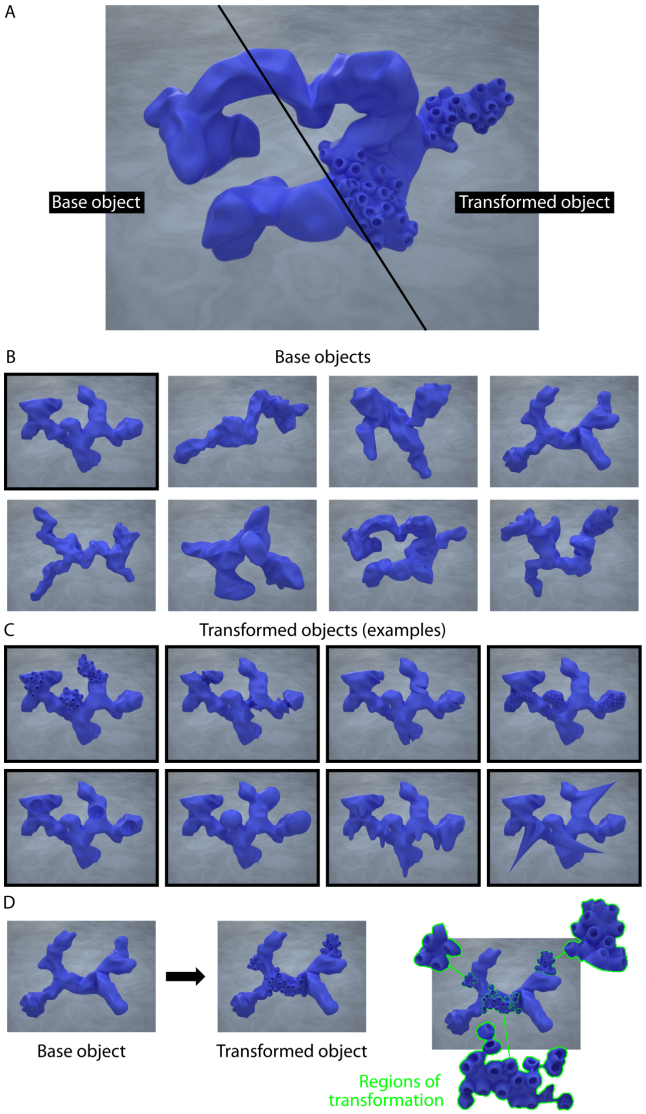

Figure 3. (A) Example of a contrasted base and transformed object. (B) The eight base objects. (C) The eight different versions of the first base object (black frame in A) with localized transformations. We used the same tools to produce eight versions of each of the other base objects. (D) Another example for a base object and transformed object, with a depiction of the transformed object where we highlight and magnify the regions of transformation.

2.1.3 Procedure. In each trial, we presented a participant with all eight stimuli of one transformation class at once (e.g., all stimuli shown in Fig. 1B)–arranged at random across two monitors, with four stimuli each. We asked participants to "name or describe the change or process that happened to the objects". They were allowed to provide multiple responses per class, but at least one response was required. The experiment was conducted in German and responses were written on paper. The height and width of each stimulus on screen was 16.0 × 21.3 cm (about 18.33 × 24.41° of visual angle at a monitor distance of about 50 cm). Stimuli were presented on a white background on two Dell U2412M monitors at a resolution of 1920 x 1200 pixels. Each participant responded to transformation classes in random order. Participants were never shown the base objects.

2.1.4 Analysis. Prior to data collection, we decided to exclude nouns referring to object features (e.g., "holes") or descriptions referring to materials (e.g., "stone") from the analysis. This eliminated 21% of total responses, with 13% due to participants responding "holes", "balls, and "spikes" (to transformations in Fig. 5A, B, and D, respectively). While these responses could plausibly be a means by which observers sought to describe causal processes, we cannot rule out that they simply describe visible shape features. Note that this percentage of exclusions refers to the entirety of all responses, including multiple responses from single participants. Only a single combination of participant and transformation class (out of 48, i.e. 2%) was completely removed. Data can be obtained from https://doi.org/10.5281/zenodo.2609016.

2.2 Results and Discussion

Responses were tightly constrained, and showed high mutual agreement between participants. Even though no restrictions were placed on the participants' responses, the average agreement across transformations for the single most frequent response was about 78% (range: 44-100%) (Figs. 4 and 5). For five out of eight transformation classes, this agreement was > 80% (Fig. 4A, C, D; Fig. 5A, C). Additionally, for all but one transformation (Fig. 4B), most participants agreed on a single description fitting the stimuli best (as can be seen from the difference in word size in the word clouds). We replicated these findings in a control experiment where each participant was presented with just a single stimulus (supplementary Figure S1).

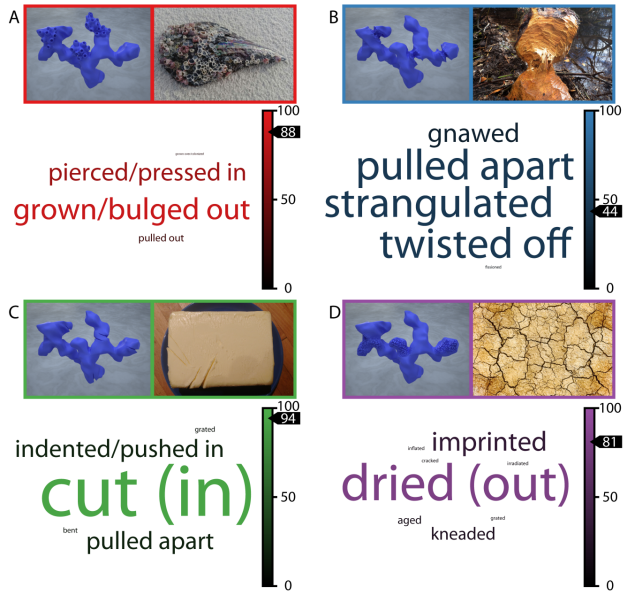

Figure 4. Results of the free naming task for first four of the transformation classes. Each of the panels (A)-(D) shows an example stimulus from our experiment, a real-world example that was chosen by us to express the same or similar transformation, and participant responses as a word cloud. The size of each word within a cloud depends on its frequency across all participants (e.g., in (A) "grown/bulged out" was more frequent compared to "pierced/pressed in"). The color of each word within a cloud depends on the correspondence between participants–the color bar gives correspondence between % of participants–with the value of highest correspondence marked in the color bar (e.g., in (A) "grown/bulged out" was produced by 88% of participants). All images are reprinted with permission: 'Overgrown seashell' and 'Cut butter' by first author (A, C); 'Gnawed wood' by Lamiot, 2014, is released under Create Commons CC BY-SA 3.0 [https://creativecommons.org/licenses/by-sa/3.0/] (B); 'Cracks' is released free of copyrights under Creative Commons CC0 (D).

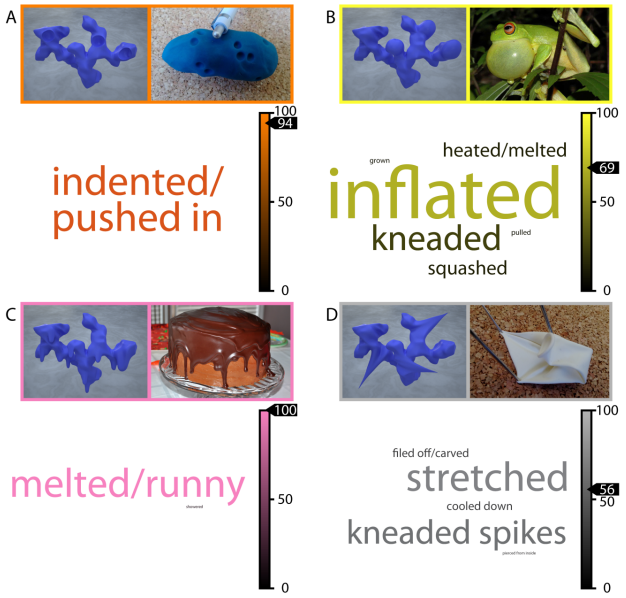

Figure 5. Results of the free naming task for second four of the transformation classes. For details see Fig. 3. All images are reprinted with permission: 'Indented putty' and 'Stretched balloon' by first author (A, D); 'Litoria chloris calling' by Froggydarb, 2006, is released under Create Commons CC BY-SA 3.0 [https://creativecommons.org/licenses/by-sa/3.0/] (B); 'Chocolate cake' by The Bake More, 2010 (C).

In the following, we refer to the transformation classes with the names provided by participants (Figs. 4 and 5): "grown", "twisted", "cut", "dried", "indented", "inflated", "melted", and "stretched".

3 Experiment 2 (Classification task)

Having validated our stimuli, we tested whether participants could separate a shape into the different contributions of base object and transformation (i.e., causal history).

3.1 Materials and Methods

3.1.1 Participants. A new group of 15 students from Justus-Liebig-University Giessen, Germany, with normal or corrected vision participated in the experiment for financial compensation. Participant procedures were the same as in Experiment 1.

3.1.2 Stimuli. Stimuli were the same 64 objects as in Experiment 1, printed out on laminated cards (Fig. 6A). The height and width of each stimulus was 13.0 × 17.3 cm. Viewing distance was unconstrained. All stimuli can be obtained from https://doi.org/10.5281/zenodo.2609016.

3.1.3 Procedure. In each experimental trial, we presented a participant with all 64 stimuli at once, laid out on a table in a random arrangement with all stimuli visible (Fig. 6A). In the *transformation instruction* condition, we asked participants to "group objects according to what happened to them". We told them they could form as many groups, with as many members per group as

they wanted to. After they were finished, we asked them to leave the room and the experimenter registered their responses. In the *shape instruction* condition, we asked the participants to "group objects according to the shape they had before something happened to them". To reduce the frequency of unintentional errors, we asked them after they finished each condition to go through their group arrangements once more and make sure that every object was in the group they thought it belonged to. Each participant responded to each of the two instructions in random order. Note that participants were never shown the base objects.

3.1.4 Analysis. We organized responses in grouping matrices, where cells were defined by combinations of all 64 stimuli, and each cell was given a value depending on whether the two stimuli were grouped by participants or not (Fig. 6B). This allows us to compare responses to predictions for groupings based on ground truth transformation or ground truth base object, respectively. If participants were able to scission shape, they should group together all stimuli from one transformation class in the transformation instruction and all stimuli derived from the same base object in the shape instruction. Data can be obtained from https://doi.org/10.5281/zenodo.2609016.

3.2 Results and Discussion

Participant's performance was close to optimal, indicated by very high correlations between responses and prediction matrices for both instructions (Fig. 6B; transformation instruction: $R^2 = 0.99$. Note that identical stimuli were presented in the two conditions. The differences in responses must reflect the participants' ability to group according to one criterion or the other at will. These results strongly suggest that for these stimuli participants are excellent at teasing apart the original shape features from those due to the transformations applied to the objects.

Some caution should be taken in interpreting these results as evidence for 'shape scission' in its fullest form (i.e., segmenting a single unfamiliar shape into distinct, superimposed causes). In particular, here, rather than showing isolated shapes, we presented many shapes that were related to one another in terms of the original shape or transformation. Thus, we cannot infer how individual shapes would be interpreted when presented in isolation. It is likely that participants used comparisons between the stimuli to identify the factors 'original shape' and 'transformation'. Nevertheless, this factorization is itself a non-trivial ability. Due to the stimulus design, shapes could not be template-matched to one another. Instead, participants must somehow identify and abstract the statistical features that are common to different shapes that have been subjected to the same transformation. Indeed, the fact that they could at will select different internal criteria depending on the instructions—presumably based on different shape features—suggests that sophisticated perceptual organization processes underlie the categorization decisions.

It has been shown previously that transformations are also inferred from changes in 2D contour (e.g., Chen & Scholl, 2016; Pinna, 2010; Spröte et al., 2016). Moreover, 2D contours from projections of 3D objects define local shape at the corresponding positions on the 3D object (e.g., Norman, Dawson, & Raines, 2000; Richards, Koenderink, & Hoffman, 1987). This suggests that for many transformations, contour information alone might be sufficient to classify objects according to their transformations. For example, "melted" and "stretched" objects from our stimulus set will have clearly distinct 2D contours. However, for other inferences such as the localization or magnitude of transformation (Experiments 3 and 4), we have to rely on 3D object representations.

4 Experiment 3 (Location task)

After showing that participants assign meaning to the transformations in our stimuli (Exp. 1) and are able to group stimuli according to their base object and the applied transformation, (Exp. 2), we sought to test whether they localize features introduced into the object by shape-altering transformations. Specifically, we asked them to mark the regions affected by a transformation, by digitally painting directly onto the stimuli.

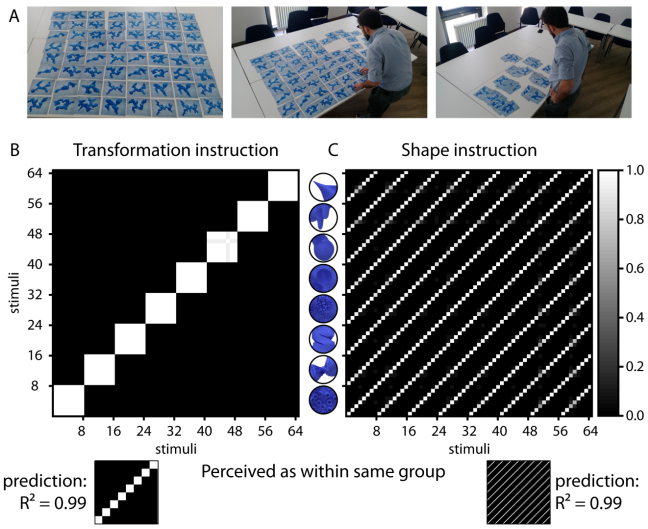

Figure 6. Procedure and results of the classification task. (A) A mock participant completing the task according to the transformation instruction (left: initial random arrangement; middle: during task; right: final arrangement). (B) Grouping matrices for transformation instruction (left) and shape instruction (right). The matrices show grouping responses averaged across all participants, with individual response = 1 (white) when two stimuli were in the same group and response = 0 (black) when two stimuli were in different groups. Stimuli are arranged along the two axes so that those from the same transformation class are at consecutive positions, with the same order of base objects within each class. The resulting prediction matrices are plotted small below the result matrices. R^2 values are based on the correlations between the (lower triangular parts of) prediction and response matrices.

4.1 Materials and Methods

4.1.1 Participants. A new group of 56 students from Justus-Liebig-University Giessen, Germany, with normal or corrected vision participated in the experiment for financial compensation. Participant procedures were the same as in Experiment 1.

4.1.2 Stimuli. Stimuli were the same 64 objects as in Experiment 1, presented via an iPad. The height and width of each stimulus on screen was 15.5 × 20.7 cm (free viewing distance). Stimuli were presented at a resolution of 2732 x 2048 pixels. All stimuli can be obtained from https://doi.org/10.5281/zenodo.2609016.

4.1.3 Procedure. In each trial, we presented one of the 64 stimuli on an iPad and asked the participant to "paint the areas on the object that appear deformed to you" (Fig. 7A). We also told participants that painted areas outside the object contour would not be considered in the analysis. Responses were made via an Apple Pencil. The interface allowed them to change the size of the virtual brush and to remove erroneous strokes with an erase tool. Once the participant was satisfied that they had painted all the relevant portions of the stimulus, they proceeded to the next trial by tapping a button on the interface. At the end of each trial, responses for each pixel in each stimulus were binary (either painted or not). Each participant responded to 16 stimuli in random order, with each transformation and base object occurring twice within the 16 trials. After all participants had completed the experiment, each stimulus was responded to by 14 participants. Note that participants were never shown the base objects.

4.1.4 Analysis. First, all painted regions outside each object were discarded. Second, we averaged the binary response maps across all 14 participants for each stimulus, and normalized to the range [0,1]. This produced a single image per stimulus, in which lighter regions indicated more frequently painted regions across participants (Fig. 7C). This allowed us to compare responses to predictions for transformation location groupings based on ground truth transformation. Ground truth was obtained from the differences between the 3-dimensional meshes of the transformed object and the base object. Specifically, we computed the Euclidean distance in 3D between each visible vertex in the transformed object and its counterpart in the base object, and then projected this with the same view frustum as the stimulus renderings, such that for every pixel in the stimulus, there was a corresponding value of the difference between transformed and original shape. We normalized all differences per shape to the range [0,1] and plotted it, producing a single image per stimulus, in which lighter regions indicated larger differences between transformed and base

object (Fig. 7B). In the following, we correlated the response image to the ground truth deformation image across all stimuli of each transformation class. For comparison, we calculated chance level correlation by comparing the response image for a given stimulus with the mesh difference predictions from all but the actual transformation class for the same stimulus. For example, we correlated the response to the "cut" base object A to the ground truth deformation images of the "inflated" base object A, the "dried" base object A, and so on, across all stimuli per transformation class. We calculated *z* scores to compare the average correlation coefficient across stimuli within each transformation class against chance level for that class: if participants were able to localize the transformations, correlations should be higher than chance. Note that *z* scores were computed based on pixel-by-pixel correlations, thereby presumably somewhat overestimating significance as participants were not able to independently set individual pixel values. Data can be obtained from https://doi.org/10.5281/zenodo.2609016.

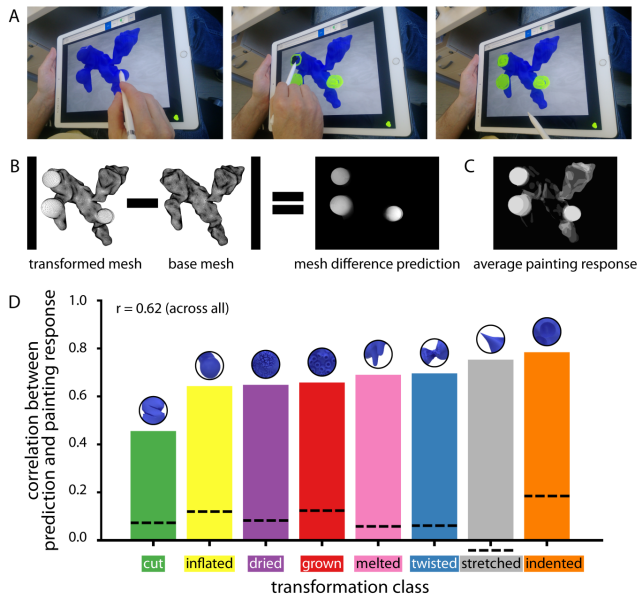

Figure 7. Procedure, analysis and results of the painting task. (A) A mock participant is completing a single trial of the task. (B) For analysis, we calculated the difference between the transformed mesh and the mesh of the base object. This process is illustrated here with an example shape and transformation. Then, we calculated the resulting mesh difference prediction with the (C) average painting response of 14 participants for this shape. (D) Correlation between mesh difference prediction and average painting response across all stimuli per transformation class. The dotted lines indicate chance correlation, obtained by correlating the average painting response per stimulus with the mesh difference predictions from all but the actual transformation class for the same stimulus, across all stimuli per transformation class.

4.2 Results and Discussion

We found substantial correlations between perceived and ground truth regions of transformation on the objects for all transformation classes, varying in range between r = 0.45 and r = 0.78 (Fig. 7D). All of the correlations were clearly above chance, with all $z > 402.49$, all $p < .001$. The relatively low correlation for the "cut" transformation can be explained by the very localized effects of this transformation, together with a relatively coarse painting style most participants adopted (even though they could adjust the size of their virtual brush, hardly anyone chose to do so). The results show that even though participants were never shown the non-transformed original shape, they could identify and localize features that were introduced by a transformation that had been applied to it.

5 Experiment 4 (Deformation rating)

5.1 Materials and Methods

5.1.1 Participants. A new group of 15 students from Justus-Liebig-University Giessen, Germany, with normal or corrected vision participated in the experiment for financial compensation. Participant procedures were the same as in Experiment 1.

5.1.2 Stimuli. For Experiment 4, we added three intermediate levels of transformation by using Blender Shape Keys (Blender 2.76, Stichting Blender Foundation, Amsterdam, NL) to interpolate between each base object and each of its transformed versions, resulting in five transformation levels (0%, 25%, 50%, 75%, 100% transformation). We rendered these new stimuli in the same way as the previous ones, obtaining a total of 320 stimuli (5 magnitude levels × 8 transformations × 8 base objects).

As the 0% condition is identical across all transformations (i.e., it is the non-transformed original shape), we removed the (7 transformations × 8 base objects) = 56 redundant stimuli, yielding a total of 264 unique stimuli. All stimuli can be obtained from https://doi.org/10.5281/zenodo.2609016.

5.1.3 Procedure. In each experimental trial, we showed participants one of the 264 stimuli and asked them to "make a rating on the provided scale between *not deformed* and *strongly deformed*". After choosing a position on the rating bar they proceeded to the next trial by clicking. Participants did not see any of the stimuli before the first trial. The height and width of each stimulus on screen was 22.5 × 30.0 cm (about 25.78 × 34.38° of visual angle at a monitor distance of about 50 cm). Stimuli were presented on a gray background on a Dell U2412M monitor at a resolution of 1920 x 1200 pixels. Each participant responded to all 264 stimuli in random order.

5.1.4 Analysis. For analysis, we derived two mesh measures. As a first measure, we calculated the ground truth *magnitude of deformation*—equivalently to Experiment 3—by calculating the 3D Euclidean distance between each visible vertex in the transformed object and its counterpart in the base object (Fig. 8A,B) and rendering the result into an image. We normalized all differences across all shapes and transformations to the range [0,1], and averaged these vertex differences per shape, yielding a single value of deformation per stimulus. This measure was inspired by our previous research in which we showed that perceived deformation was a very good predictor of perceived softness–and strongly correlated with actual mesh deformation (Paulun, Schmidt, van Assen, & Fleming, 2017). We reasoned that participants in the current study might also rely on deformation to estimate the magnitude of object change. However, in contrast to our previous studies, changes in object shape were local rather than global (Paulun et al., 2017; Schmidt & Fleming, 2016; Spröte & Fleming, 2016). As results of Experiment 3 show that participants can infer the spatial extent of transformations, we decided to compute, as a second measure, the ground truth *area of deformation* by calculating the area of deformed faces (i.e., faces with at least one vertex with difference > 0; again only considering faces that were visible from the point of view) (Fig. 8C). We averaged this area of deformed faces per shape, and normalized all areas across all shapes and transformations to the range [0,1], yielding a single value of deformation area per stimulus. We performed multiple regression analyses across all stimuli and within transformation classes to measure the predictive power of magnitude and area of deformation for participant responses. If responses can be explained by these two factors, this suggests that participants might use perceptual equivalents of these (i.e., perceptual quantities that correlate with these factors) to infer the magnitude of transformation. Data can be obtained from https://doi.org/10.5281/zenodo.2609016.

5.2 Results and Discussion

The average perceived deformation increased across the five transformation levels within each transformation, with overall perceived deformation varying between transformations (Fig. 9A). For analysis, we correlated participants' average responses per stimulus with the magnitude and area of deformation (Fig. 9B). Even across all stimuli there is a substantial correlation between perceived deformation and the ground truth magnitude of deformation as well as the area of deformation (the correlations between both mesh measures and transformation levels were about $R^2 = 0.41$ and 0.46, respectively). As correlation between the two measures is relatively weak in our set of stimuli (Fig. 9C), we entered them both as predictors in a multiple regression analysis.

A multiple regression across all transformation classes explained $R^2 = 0.68$ of the variance, with regression weights of 0.43 and 0.56 for magnitude and area of deformation, respectively. Based on previous work on the estimation of high-level object and material properties (Long, Yu, & Konkle, 2017; Schmid & Doerschner, 2018; Schmidt, Hegele, & Fleming, 2017; van Assen, Barla, & Fleming, 2018), we hypothesized that

participants might use different mid-level features of shape (such as ripples, bumps, or concavities and holes), depending on the type of transformation. This would predict that responses to different transformations are explained by the two different measures to a different degree. To test this hypothesis, we also performed multiple regressions within each transformation class. In accordance with our hypothesis, the overall explained variance increased to R^2 = 0.82, with regression weights showing distinct patterns between transformations (Fig. 10; for scatter plots see supplementary Figures S2 and S3): while for some transformation classes, perceived deformation was almost exclusively explained by the magnitude of deformation, for other transformation classes, perceived deformation was explained to a substantial amount by the area of deformation.

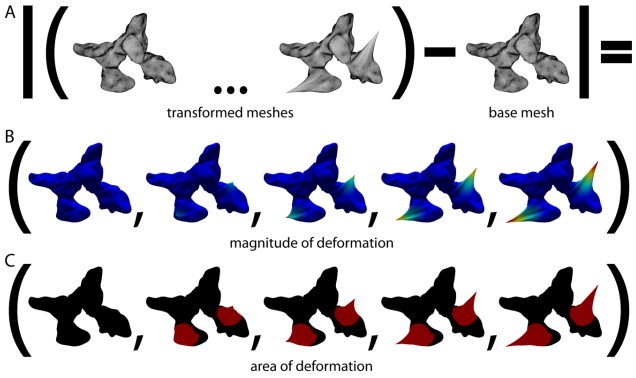

Figure 8. Calculation of magnitude and area of deformation measures illustrated with an example shape and transformation. (A) Equivalently to analysis of Experiment 3, we calculated the difference between the transformed meshes of each deformation level and the mesh of the base object. Thus, we obtain for each stimulus a unique value of (B) its magnitude of deformation, defined by the average distance of each deformed vertex from non-deformed vertex location, and a value of (C) its area of deformation, given by the average area of deformed faces.

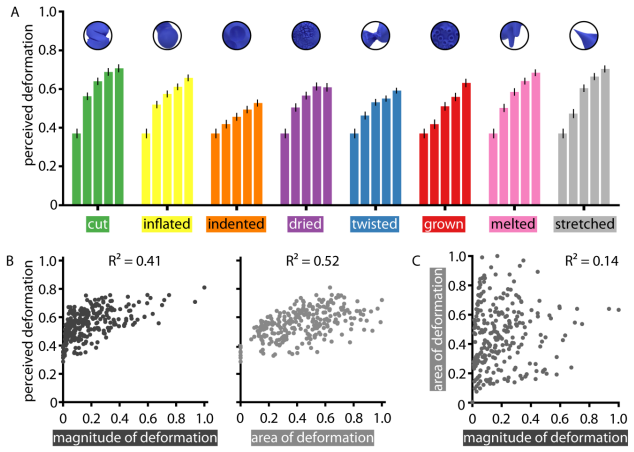

Figure 9. (A) Average ratings of perceived deformation for the five different transformation levels (bars from left to right), separately for each transformation. (B) Scatter plots for average ratings of perceived deformation as a function of magnitude of deformation (left) and area of deformation (right). (C) Scatter plots for the relationship between the magnitude of deformation and the area of deformation. In all plots (B,C), each data point represents a single stimulus; magnitude and area of deformation were normalized to the range [0,1]; and R^2 values are based on the Pearson correlation coefficient.

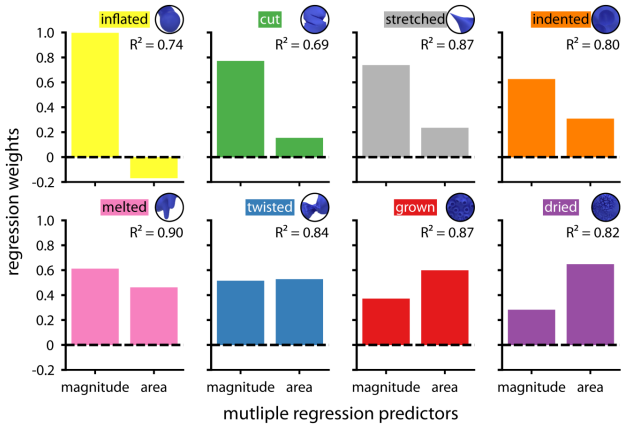

Figure 10. Regression weights from multiple linear regression per transformation class on average deformation ratings, using magnitude of deformation and area of deformation as predictors. Plots are ordered from left to right and top to bottom with decreasing weights for magnitude of deformation, and increasing weights for area of deformation.

6 General Discussion

6.1 Summary of Findings

Transformations of object shape are ubiquitous in our visual environment. Here, we document the capability of our visual and cognitive systems to solve two

complementary and linked inferences: recognizing objects across transformations, and recognizing transformations across objects. Specifically, we tested whether we can distinguish the causal origin of different features, teasing apart the characteristics of the original shape from those imposed by transformations. In four experiments, we demonstrate that participants (i) show high levels of mutual agreement in naming transformations, (ii) can classify objects according to their original shape or according to their transformations, (iii) can mark the areas affected by the transformations, and (iv) can to some extent estimate the magnitude of transformations. This shows that observers distinguish between contributions of the previous object shape and its recent transformations. In particular, the fact that participants can sort the same stimuli almost perfectly into groups according to either their original shape or the transformation, suggests that they can, at will, attend to the different features associated with these distinct causes. We suggest that it may be useful to pose these abilities as a 'scission' of object shape into different contributions, analogous to scission problems in other areas of visual perception such as the separation of retinal luminance values into separate representational layers of illumination and reflectance.

6.2 Functional Utility of Shape Scission

Why would our visual and cognitive systems separate shapes into different causal layers? Scission models in lightness perception propose that layers of illumination and reflectance are used to achieve lightness constancy. Similarly, 'shape scission' could aid us identify and understand transformations of object shape, to facilitate object constancy—for example, by recognizing objects across diverse viewing conditions or organisms across growth (Schmidt & Fleming, 2016). The ability to infer the processes that have shaped an object could also support object classification for novel stimuli, especially when only a single or very few exemplars are presented—for example, in predicting or generating plausible variants of transformed objects (Feldman, 1992, 1997). Finally, identifying and understanding transformations is used to infer object properties, such as softness of deformable objects (e.g., Bi, Jin, Nienborg, & Xiao, 2018; Bi & Xiao, 2016; Bouman, Xiao, Battaglia, & Freeman, 2014; Paulun et al., 2017; Schmid & Doerschner, 2018; Schmidt, Paulun et al., 2017) and anticipate future object behaviors (such as bouncing or shattering; Alley, Schmid, & Doerschner, submitted; Bates, Yildirim, Tenenbaum, & Battaglia, 2015) and motor affordances (such as locations and grip and load forces in grasping, and forces in probing objects; Klein, Maiello, Paulun, & Fleming, 2018; Zöller, Lezkan, Paulun, Fleming, & Drewing, 2018). Based on the inference that an object has been stretched, we can make an educated guess about its internal properties (e.g., it is rather malleable, elastic and hollow) and potentially about its material category (e.g., it is more likely to be a rubber balloon than a leather ball). This will also guide us in predicting object behavior and in handling it: for example, a hammer with a cracked handle is unsuitable for hammering; withered salad should not be eaten; a bent knife is bad for cutting bread. Taken together, 'shape scission' provides us with a layered representation of object shape that we can use to identify and understand transformations which is potentially important for a wide range of perceptual and cognitive tasks.

6.3 Relation to previous work

Our findings are in line with previous evidence and theoretical work suggesting that we represent object shape at multiple representational layers (Green, 2015), some of which we access only when needed (Op de Beeck, Torfs, & Wagemans, 2008; Pinna, Koenderink, & van Doorn, 2015). Depending on their task, participants grouped the objects according to different features. This corroborates findings from a recent study in our lab, where participants were able to group the same images of transformed (e.g., twisted, bent) samples of different materials (e.g., cardboard, putty) either by transformation or material (Schmidt & Fleming, 2018). In related work, Pinna (2010) demonstrated that we infer particular transformations ("happenings") as well as

material properties (Pinna & Deiana, 2015) from simple, deformed 2D stimuli like rectangles or grid-like arrangements of squares. Our study extends these findings by exploring a wider range of transformations; by using more carefully controlled stimuli (all factors except original shape and transformation were held constant); providing convergent evidence for 'shape scission' from multiple tasks (such as spontaneous identification of transformations and explicit multiple groupings of related stimuli); and by computing geometrical shape quantities to predict judgments of transformation location and magnitude, which was not possible with the photographs of natural materials used in Schmidt & Fleming (2018).

A deeper question is *how* do we identify regions of transformation and assign meaning to them? In his seminal work, Leyton (1989) suggested that we infer causal history by (i) identifying symmetry axes terminating at local curvature extremes of its contour, and (ii) assuming that transformations proceed along the direction of these local symmetry axes. He also argued that the perceived order of these transformations depends on their detail: observers would assume that larger changes have taken long compared to smaller changes. Consequently, the (smaller) details of the contour should be perceived as results of the most recent transformations ("deblurring"). However, despite its intuitive appeal, empirical evidence has been scarce for the basic model (Leyton, 1986a, 1986b) as well as for the meanings that Leyton (1989) assigned to the different transformations following from his theoretical work (e.g., "protrusion" or "squashing"). Moreover, the mathematical foundations of his theory have been the subject of strong criticism (Hendrickx & Wagemans, 1999). It also remains unclear how the visual system might group multiple process-related symmetry axes of a given shape into coherent global interpretation (e.g., twisting). As a result, complex, texture-like transformations, like ours, cannot easily be explained in terms of simple, local geometric transformations as described in Leyton's framework.

6.4 Perceptual representations of shape

We suggest that observed objects are represented in a multidimensional shape feature space, which has been established based on previous visual experience (Brincat & Connor, 2004; DiCarlo & Cox, 2007; Leeds, Pyles, & Tarr, 2014). In this shape space, each object occupies a specific location, while object clusters (e.g., instantiations of a single object under different viewpoints, or objects sharing a perceptual category) are organized along manifolds within the space. These different manifolds are not mutually exclusive so that they enable parallel representations of multiple contributions to object shape. For example, all "twisted" objects, signified by parallel, spiraling creases, would be organized along a specific manifold. We speculate that the manifolds are established based on prior experiences and related assumptions about likely variations of objects (Feldman, 1992, 1997; Lowet, Firestone, & Scholl, 2018; Shepard & Cermak, 1973). Such multi-dimensional representations are well suited to inferring physical properties of objects and materials. For example, we have recently shown that when judging the viscosity of liquids, the visual system represents stimuli using a number of complementary 'mid-level' shape and motion features (Van Assen, Barla & Fleming, 2018). Such features ensure that very diverse stimuli with similar physical properties (e.g., the same liquid interacting with very different scenes) are brought close together in the representational space, making it easier for high-level processes to read out the viscosity. We reason that if such an approach works for complex, highly mutable deforming shapes like liquids, it is likely also works for other types of shape-altering transformation, like the ones used here. Only when observers share similar manifolds, they will name transformations across objects with high mutual agreement (Pinna, 2010; Schmidt & Fleming, 2018). An important topic for future research is to develop stimulus-computable feature spaces for representing objects with different transformations applied, so that this conjecture can be tested. It is particularly interesting to ask which specific features

define the space, and how they are acquired through learning. A successful model should be able to predict quantitative similarity relationships between stimuli (including errors as well as successes), and explain how both 'base object' and 'transformation' components can be decoded from the positions of items in the feature space. Such models would allow us to progress from general statements about the use of high-dimensional feature spaces to specific quantitative predictions.

In addition to identifying transformations, our findings show that we can also estimate properties of the transformation process, such as the location and the magnitude of the transformation. Localization may be based on detecting the features associated with a known transformation (e.g., the distinctive rough edge of a tear) of transformation shape features on the object, analogous to the detection of features associated with familiar objects (e.g., the trunk of an elephant). However, it is also interesting to ask how the visual system could segment and localize *unfamiliar* transformations, based on non-stationarities in the surface statistics (e.g., a localized increase in roughness).

The estimation of transformation magnitude is perhaps an even more complex computational problem as it involves pooling and weighing multiple features. A good deal of the variance in participants' responses was explained by a combination of (i) the mesh difference between the current and the previous object shape (magnitude of deformation) and (ii) the size of the surface area affected by the transformation (area of deformation). Although we do not think that the visual system computes exactly these measures (not least as they would require an explicit estimate of the original shape), it may use heuristics that approximate these quantities. For example, to estimate the magnitude of deformation the visual system might rely on the heuristics related to amodal completion, particularly when features of the original object are "occluded" by the transformation-related features, but the surrounding surface structure remains intact. Under these circumstances continuation based on local contours and surface structure, or global factors such as symmetry (e.g., Tse, 1999) could facilitate recovery of the original shape 'hidden' by the transformation. The 'amodally completed' shape could be compared to the observed shape to estimate the area and magnitude of transformation.

Notably, the relative weights of the two mesh measures were different between transformations. For example, the area of deformation explained responses best in rather "textural" surface transformations (such as "grown" and "dried"). This suggests that different transformations are not only characterized by their different (combinations) of shape features but that we might also use different (combinations) of heuristics to reveal transformation attributes (e.g., transformation magnitude).

6.5 Limitations

Our findings and conclusions are limited by the selection of base objects and transformations, and do not generalize to all other objects and transformations. Indeed, in a previous study with more global transformations, we found that observers were only moderately good at inferring causal history (Schmidt, Paulun et al., 2017). There are many instances where observers would not, even in principle, be able to infer the causal history of objects. Not all transformations involved in the generative processes producing an object (such as manufacture, biological growth or self-organization) leave traces in the shape of the object. For example, we are not able to infer all processes and forces involved in the production of a computer mouse. Nor can we infer transformations introducing shape features that are indistinguishable from those of the original object. Indeed, for familiar objects, transformations are presumably only perceived as such when they represent a deviation from the typical shape characteristics of the class: the shape of a bumper bar is not seen as due to transformations, but the dents from the latest parking mishap are (also see discussion in Schmidt & Fleming, 2018).

This also demonstrates that depending on the specific stimulus and task, the inferences span both perceptual and cognitive abilities, with different levels of

detail in the inference about the sequence of events that led to the observed state of the object (see also discussion in Spröte & Fleming, 2016). While the distinction between perception and cognition is important (Chen & Scholl, 2016), many tasks naturally span both—with 'shape scission' tasks being of this type. For example, the consistency of naming responses across participants is illustrating cognitive, semantic judgments based on shared previous experience. By contrast, the other reported tasks do also involve perceptual components. Especially the ability of participants to localize and estimate the magnitude of transformations suggests that they can perceptually parse objects into transformed and non-transformed regions. The correlation of magnitude estimations with objective mesh measures (i.e., characteristics of the physical object shape) emphasizes the role of visual (shape) processing in making those inferences.

Finally, our selection of transformations was ad-hoc and–because they were direct mesh deformations–not subject to the physical constraints of real-world objects and materials (see discussion in Schmidt & Fleming, 2018). Also, our findings were obtained under close to ideal conditions for shape perception: objects were static, with hardly any scene context, no interreflections, no deep shadows and so on. Real-world shape perception is subject to these and other sources of noise which we eliminated by careful rendering. Nevertheless, our findings illustrate some salient examples in which we perceptually 'understand' shape by parsing and interpreting causally significant features.

6.6 Conclusions and future directions

In a series of experiments, we demonstrated the ability of participants to identify and understand object transformations, and to estimate properties of the transformation process. In particular, they could tease apart the shape features of the original object from those introduced by the transformation. We suggest that one useful way of framing such inferences is in terms of 'shape scission'—the separation of object shape into different representational layers of intrinsic (such as typical shape, object material) and extrinsic (such as orientation, lighting conditions, causal history) contributions to the observed shape. Although caution is required in interpreting our results as the purest form of 'shape scission', participants do seem capable of distinguishing and comparing different kinds of features within a given object.

Many questions remain. Which perceptual information do we use to separate original and transformed regions of the objects? How do we decide which regions belong to the original object and which belong to the transformation? How detailed is our representation of the original object, and under which circumstances can we 'invert' the transformations to recover its original features? What computations recover the causal history of objects from the feature representations (i.e., how do we assign 'meaning' to observed features)? What are the neural representations of these different types of perceptual and cognitive processes (Op de Beeck et al., 2008; Ward et al., 2018)? Computational and neuroimaging techniques should be combined to answer these and related questions.

Acknowledgments

We wish to thank Jasmin Kleis, Hendrik Will and Marcel Schepko for data collection and Yi-Chia Chen for helpful comments.

Funding

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)–project number 222641018–SFB/TRR 135 TP C1 and by the European Research Council (ERC) Consolidator Award 'SHAPE'–project number ERC-CoG-2015-682859.

Competing interests

The authors declare no competing interests.

References

Adelson, E. H. (2000). Lightness perception and lightness illusions. In M. Gazzaniga (Ed.), *The New Cognitive Neurosciences* (pp. 339–351). Cambridge, MA: MIT Press.

Alley, L. M., Schmid, A. C., & Doerschner, K. (submitted). Neo's spoon and Newton's apples: Prediction of rigid and non-rigid deformations as cues to material properties.

Anderson, B. L. (1997). A theory of illusory lightness and transparency in monocular and binocular images: The role of contour junctions. *Perception*, *26*(4), 419–453. https://doi.org/10.1068/p260419

Anderson, B. L. (2003). The role of occlusion in the perception of depth, lightness, and opacity. *Psychological Review*, *110*(4), 785–801. https://doi.org/10.1037/0033-295X.110.4.785

Anderson, B. L., & Winawer, J. (2005). Image segmentation and lightness perception. *Nature*, *434*(7029), 79–83. https://doi.org/10.1038/nature03271

Arnheim, R. (1974). *Art and visual perception: A psychology of the creative eye*: Berkeley, CA: University of California Press.

Barrow, H. G., & Tenenbaum, J. M. (1978). Recovering intrinsic scene characteristics. In A. Hanson & E. Riseman (Eds.), *Computer Vision Systems* (pp. 3–26). New York: Academic Press.

Bates, C. J., Yildirim, I., Tenenbaum, J. B., & Battaglia, P. W. (2015). Humans predict liquid dynamics using probabilistic simulation. In: Proceedings of the 37th Annual Conference of the Cognitive Science Society, Pasadena, CA, July 22–25, 2015. Cognitive Science Society., 172–177.

Bi, W., Jin, P., Nienborg, H., & Xiao, B. (2018). Estimating mechanical properties of cloth from videos using dense motion trajectories: Human psychophysics and machine learning. *Journal of Vision*, *18*(5), 12. https://doi.org/10.1167/18.5.12

Bi, W., & Xiao, B. (2016). Perceptual constancy of mechanical properties of cloth under variation of external forces. In E. Jain & S. Joerg (Eds.), *Proceedings of the ACM Symposium on Applied Perception - SAP '16* (pp. 19–23). New York, New York, USA: ACM Press. https://doi.org/10.1145/2931002.2931016

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, *94*(2), 115–117. https://doi.org/10.1037/0033-295X.94.2.115

Bouman, K. L., Xiao, B., Battaglia, P., & Freeman, W. T. (2014). Estimating the Material Properties of Fabric from Video. In *2013 IEEE International Conference on Computer Vision: 1-8 December 2013* (pp. 1984–1991). New York: IEEE. https://doi.org/10.1109/ICCV.2013.455

Brincat, S. L., & Connor, C. E. (2004). Underlying principles of visual shape selectivity in posterior inferotemporal cortex. *Nature Neuroscience*, *7*(8), 880–886. https://doi.org/10.1038/nn1278

Chen, Y.-C., & Scholl, B. J. (2016). The Perception of History: Seeing Causal History in Static Shapes Induces Illusory Motion Perception. *Psychological Science*, *27*(6), 923–930. https://doi.org/10.1177/0956797616628525

Cornelis, E. V. K., van Doorn, A. J., & Wagemans, J. (2009). The effects of mirror reflections and planar rotations of pictures on the shape percept of the depicted object. *Perception*, *38*(10), 1439–1466. https://doi.org/10.1068/p6101

DiCarlo, J. J., & Cox, D. D. (2007). Untangling invariant object recognition. *Trends in Cognitive Sciences*, *11*(8), 333–341. https://doi.org/10.1016/j.tics.2007.06.010

DiCarlo, J. J., Zoccolan, D., & Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron*, *73*(3), 415–434. https://doi.org/10.1016/j.neuron.2012.01.0

10

Feldman, J. (1992). Constructing perceptual categories. In *Computer vision and pattern recognition: Conference : Papers* (pp. 244–250). IEEE Comput. Soc. Press. https://doi.org/10.1109/CVPR.1992.223268

Feldman, J. (1997). The Structure of Perceptual Categories. *Journal of Mathematical Psychology*, *41*(2), 145–170. https://doi.org/10.1006/jmps.1997.1154

Green, E. J. (2015). A Layered View of Shape Perception. *The British Journal for the Philosophy of Science*, *28*, axv042. https://doi.org/10.1093/bjps/axv042

Hendrickx, M., & Wagemans, J. (1999). A Critique of Leyton's Theory of Perception and Cognition. Review of Symmetry, Causality, Mind, by Michael Leyton. *Journal of Mathematical Psychology*, *43*(2), 314–345. https://doi.org/10.1006/jmps.1998.1232

Kanizsa, G. (1955). Condizioni ed effetti della trasparenza fenomenica. *Rivista Di Psicologia*, *49*, 3–18. (Translation reprinted as “Phenomenal transparency”. In G. Kanizsa, (1979). Organization in vision: Essays on Gestalt perception (pp. 151-169). New York, NY: Praeger Publishers).

Klein, L. K., Maiello, G., Paulun, V. C., & Fleming, R. W. (2018). How humans grasp three-dimensional objects. *bioRxiv*, 476176.

Koenderink, J. J., van Doorn, A. J., Kappers, A. M. L., & Todd, J. T. (1997). The visual contour in depth. *Perception & Psychophysics*, *59*(6), 828–838. https://doi.org/10.3758/BF03205501

Koffka, K. (1935/1965). *Principles of Gestalt Psychology*. New York: Harcourt, Brace & World.

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. (1986a). A theory of information structure I. General principles. *Journal of Mathematical Psychology*, *30*(2), 103–160. https://doi.org/10.1016/0022-2496(86)90011-8

Leyton, M. (1986b). A theory of information structure II. A theory of perceptual organization. *Journal of Mathematical Psychology*, *30*(3), 257–305. https://doi.org/10.1016/0022-2496(86)90033-7

Leyton, M. (1989). Inferring Causal History froms Shape. *Cognitive Science*, *13*(3), 357–387. https://doi.org/10.1207/s15516709cog1303_2

Logothetis, N. K., & Sheinberg, D. L. (1996). Visual Object Recognition. *Annual Review of Neuroscience*, *19*(1), 577–621.

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

Lowet, A. S., Firestone, C., & Scholl, B. J. (2018). Seeing structure: Shape skeletons modulate perceived similarity. *Attention, Perception & Psychophysics*, *80*(5), 1278–1289. https://doi.org/10.3758/s13414-017-1457-8

Mark, L. S., & Todd, J. T. (1985). Describing perceptual information about human growth in terms of geometric invariants. *Perception & Psychophysics*, *37*(3), 249–256.

Masin, S. C. (1999). Color scission and phenomenal transparency. *Perceptual and Motor Skills*, *89*(3 Pt 1), 815–823. https://doi.org/10.2466/pms.1999.89.3.815

Metelli, F. (1970). An algebraic development of the theory of perceptual transparency. *Ergonomics*, *13*(1), 59–66. https://doi.org/10.1080/00140137008931118

Metelli, F. (1985). Stimulation and perception of transparency. *Psychological Research*, *47*, 185–202.

Norman, J. F., Dawson, T. E., & Raines, S. R.

(2000). The perception and recognition of natural object shape from deforming and static shadows. *Perception*, *29*, 135-148.

Ons, B., & Wagemans, J. (2012). Generalization of visual shapes by flexible and simple rules. *Seeing and Perceiving*, *25*(3-4), 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. *The Journal of Neuroscience : the Official Journal of the Society for Neuroscience*, *28*(40), 10111–10123. https://doi.org/10.1523/JNEUROSCI.2511-08.2008

Pasupathy, A., El-Shamayleh, Y., & Popovkina, D. V. (2018). Visual shape and object perception. In S. Murray Sherman (Ed.), *Oxford Research Encyclopedia of Neuroscience* (pp. 1–28). https://doi.org/10.1093/acrefore/9780190264086.013.75

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. https://doi.org/10.1167/17.1.20

Phillips, F., Todd, J. T., Koenderink, J. J., & Kappers, A. M. L. (1997). Perceptual localization of surface position. *Journal of Experimental Psychology: Human Perception and Performance*, *23*(5), 1481–1492. https://doi.org/10.1037/0096-1523.23.5.1481

Phillips, F., Todd, J. T., Koenderink, J. J., & Kappers, A. M. L. (2003). Perceptual representation of visible surfaces. *Perception & Psychophysics*, *65*(5), 747–762. https://doi.org/10.3758/BF03194811

Pinna, B. (2010). New Gestalt principles of perceptual organization: An extension from grouping to shape and meaning. *Gestalt Theory*, *32*(1), 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

Pinna, B., Koenderink, J., & van Doorn, A. (2015). The Phenomenology of the Invisible: From Visual Syntax to “Shape from Shapes”. *Philosophia Scientae*. (19-3), 127–151. https://doi.org/10.4000/philosophiascientiae.1135

Pittenger, J. B., & Todd, J. T. (1983). Perception of growth from changes in body proportions. *Journal of Experimental Psychology: Human Perception and Performance*, *9*(6), 945–954. https://doi.org/10.1037/0096-1523.9.6.945

Richards, W. A., Koenderink, J. J., & Hoffman, D. D. (1987). Inferring three-dimensional shapes from two-dimensional silhouettes. *JOSA A*, *4*, 1168-1175.

Riesenhuber, M., & Poggio, T. (2002). Neural mechanisms of object recognition. *Current Opinion in Neurobiology*, *12*(2), 162–168.

Schmid, A. C., & Doerschner, K. (2018). Shatter and splatter: The contribution of mechanical and optical properties to the perception of soft and hard breaking materials. *Journal of Vision*, *18*(1), 14. https://doi.org/10.1167/18.1.14

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*(8), e0202115. https://doi.org/10.1371/journal.pone.0202115

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

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. https://doi.org/10.1167/17.3.18

Schmidt, F., Spröte, P., & Fleming, R. W. (2016). Perception of shape and space across rigid transformations. *Vision Research*, *126*, 318–329. https://doi.org/10.1016/j.visres.2015.04.011

Shepard, R. N., & Cermak, G. W. (1973). Perceptual-cognitive explorations of a toroidal set of free-form stimuli. *Cognitive Psychology*, *4*(3), 351–377. https://doi.org/10.1016/0010-0285(73)90018-2

Singh, M., & Anderson, B. L. (2002). Toward a perceptual theory of transparency. *Psychological Review*, *109*(3), 492–519. https://doi.org/10.1037/0033-295X.109.3.492

Spröte, P., & Fleming, R. W. (2013). Concavities, negative parts, and the perception that shapes are complete. *Journal of Vision*, *13*(14). https://doi.org/10.1167/13.14.3

Spröte, P., & Fleming, R. W. (2016). Bent out of shape: The visual inference of non-rigid shape transformations applied to objects. *Vision Research.* Advance online publication. https://doi.org/10.1016/j.visres.2015.08.009

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

Tse, P. U. (1999). Volume completion. *Cognitive Psychology*, *39*(1), 37–68. https://doi.org/10.1006/cogp.1999.0715

Van Assen, J. J. R., Barla, P., & Fleming, R. W. (2018). Visual Features in the Perception of Liquids. *Current Biology.* Advance online publication. https://doi.org/10.1016/j.cub.2017.12.037

Ward, E. J., Isik, L., & Chun, M. M. (2018). General Transformations of Object Representations in Human Visual Cortex. *The Journal of Neuroscience : the Official Journal of the Society for Neuroscience*, *38*(40), 8526–8537. https://doi.org/10.1523/JNEUROSCI.2800-17.2018

Zöller, A. C., Lezkan, A., Paulun, V. C., Fleming, R. W., & Drewing, K. (2018). Influence of Different Types of Prior Knowledge on Haptic Exploration of Soft Objects. In Domenico Prattichizzo, Hiroyuki Shinoda, Hong Z. Tan, Emanuele Ruffaldi, & Antonio Frisoli (Eds.): *Vol. 10893-10894. LNCS sublibrary: SL3 - Information systems and applications, incl. Internet/Web, and HCI, Haptics: Science, technology, and applications : 11th International Conference, EuroHaptics 2018, Pisa, Italy, June 13-16, 2018.* Cham, Switzerland: Springer.

Supplementary Material

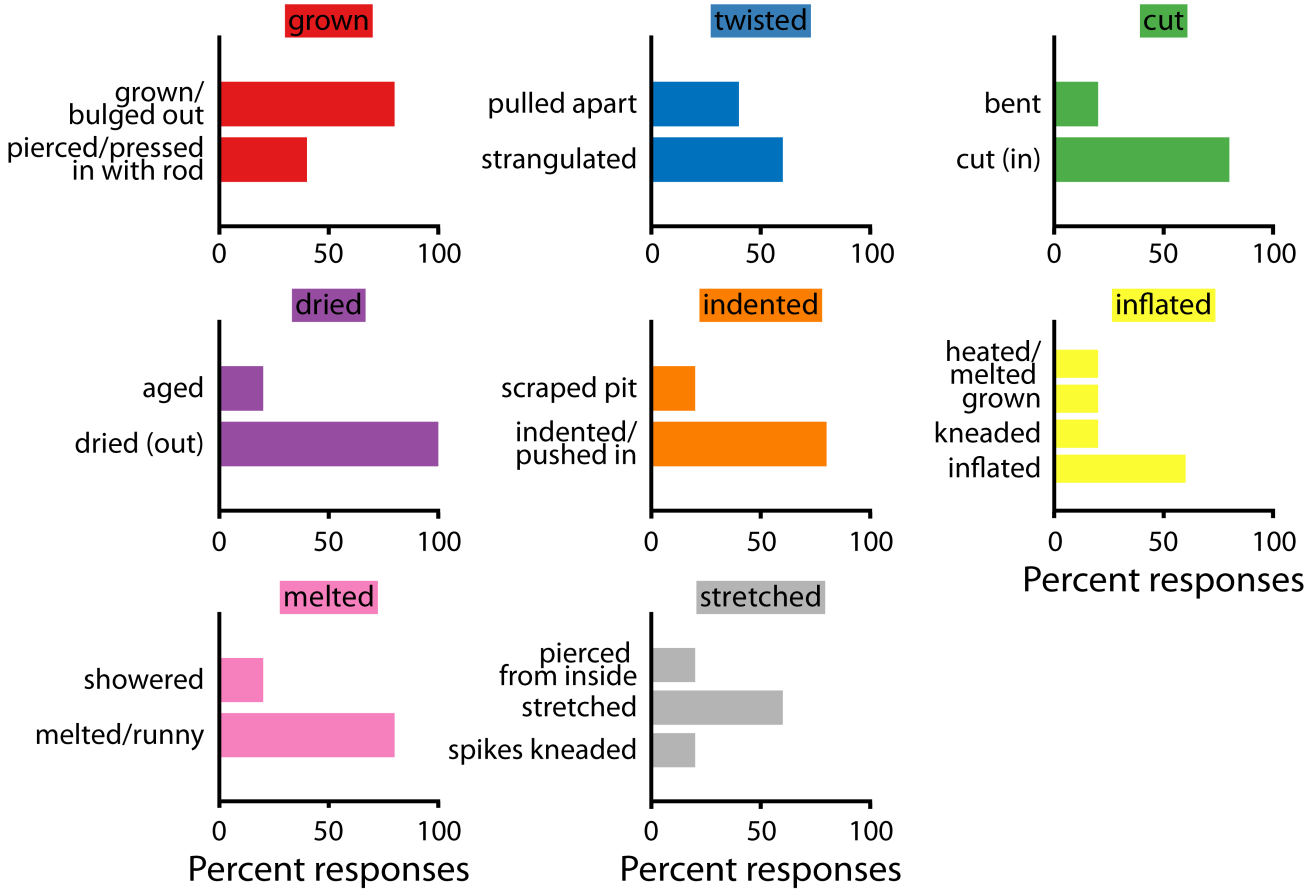

Figure S1. Replication of Experiment 1. Each participant was presented with just a single stimulus, with 5 participants judging each transformation class (on a random original shape), adding up to 40 participants (24 male, 16 female, age 29-64 years). Each graph shows the percentage of observers describing a stimulus with particular names, titles are based on results from Experiment 1.

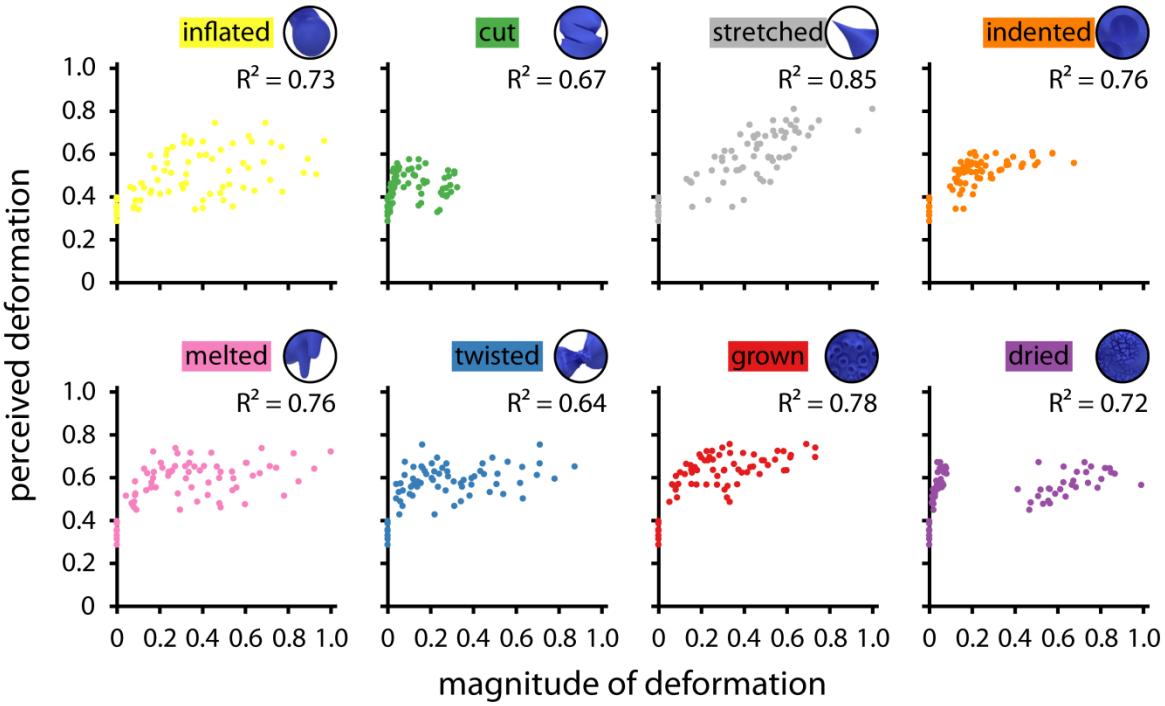

Figure S2. Scatter plots for average ratings of perceived deformation as a function of the magnitude of deformation per transformation class. Each data point represents a single stimulus. Magnitude of deformation was normalized to the range [0,1]. R^2 values are based on the Pearson correlation coefficient.

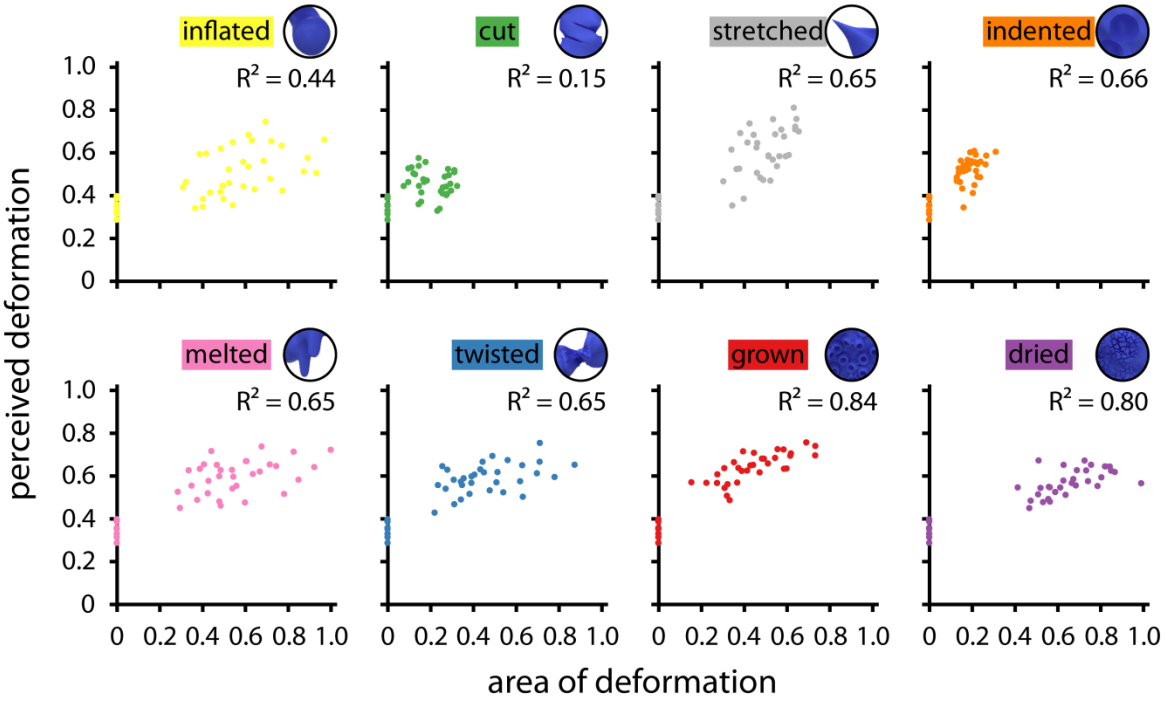

Figure S3. Scatter plots for average ratings of perceived deformation as a function of the area of deformation per transformation class. For details see Fig. S2.