Bent out of shape: The visual inference of non-rigid shape transformations applied to objects

Patrick Spröte *, Roland W. Fleming

Department of Psychology, Justus-Liebig-Universität Giessen, Germany

A B S T R A C T

In everyday life, we can often identify when an object has been subjected to some kind of transformation that alters its shape. For example, we can usually tell whether a can has been crushed, or a cookie has been bitten. Conversely, our ability to recognize objects is often robust across such shape transformations: we can still identify the can even though it has been dented. This ability to determine and discount the causal history of objects suggests the visual system may partially decompose the observed shape of an object into original (untransformed) elements plus the transformations that were applied to it. We sought to shed light on this possibility, using ‘bending’ as an example transformation. In one experiment subjects matched the degree of bending applied to random 3D shapes. We find that subjects could match the degree of bend, although there was a tendency to overestimate bends, especially for the least bent objects. In two other experiments, observers had to identify individual objects across different degrees of bending. Subjects performed significantly above chance although not as well as when the objects differed by rigid rotations without any bends (cf. traditional mental rotation experiments). Together our findings suggest that subjects can to some extent extract information about transformations applied to shapes, while ignoring other differences. At the same time subjects show a certain degree of invariance across shape transformations. This suggests scission of a shape’s representation into its causes – a base shape and transformations applied to it.

1. Introduction

Both the cookie and the croissant in Fig. 1 exhibit a concavity. However, we can easily tell that those concavities originated from quite different processes or transformations. In the case of the cookie, the concavity was created by a process of forceful removal of cookie matter—a ‘bite’—whereas the concavity in the croissant was created by shaping the dough around the concavity—a ‘bend’. The fact that we can visually understand the differences in causal history between these two quite similar shapes suggests that the visual system readily seeks to identify the generative processes that create or modify objects, as part of its visual representation of shape (Feldman & Singh, 2006; Hoffman & Richards, 1984; Leyton, 1989; Pentland, 1986a, 1986b; Richards, 1988). In this study we wanted to shed some light on the representation of transformations within the representation of whole objects. In other words, how subjects infer transformations from shape.

Inferring transformations from observations of objects is computationally challenging. Under special conditions when both non-transformed and transformed versions of an object can be observed (e.g., before and after the transformation), inferring the transformation is not trivial, but the computations required can at least be defined in a relatively straightforward way. First, the visual system would have to establish correspondence between locations on the two versions of the object, and then identify a geometrical mapping from one point set to the other (i.e., solve the ‘non-rigid registration’ problem; for a review see Crum, Hartkens, and Hill (2004)). Then, to express the complex pattern of correspondences in terms of a simple global transformation, such as a ‘twist’ or ‘bend’ of a certain magnitude, some factorization or re-parameterization of the mapping might be required. Moreover, there may be difficulties when the transformation accretes or deletes portions of the shape—as in the bitten cookie—causing unmatchable features (i.e., points that occur on one version of the object that have no counterpart on the other version, precluding correspondence).

Nevertheless, the problem of inferring transformations becomes significantly more difficult in the more typical situation, as in
One of the major problems for the visual system in 3D shape perception is the lack of three-dimensionality in the retinal image. The visual system cannot measure the third dimension directly, so it has to reconstruct it from the two dimensional retinal image. A huge amount of research has dealt with questions of how the brain uses stereopsis, motion, shading, texture, or other cues to reconstruct local 3D shape properties such as position in depth (Bülthoff & Mallot, 1988; Stevens & Brookes, 1987), surface orientation (Johnston & Passmore, 1994; Koenderink, van Doorn, & Kappers, 1992) or local curvature (Johnston & Passmore, 1994; Rogers & Cagenello, 1989). However, it should be clear that estimating local surface structure is not all there is to 3D shape perception. Even a complete and error-free map of local surface properties would leave out important information about object shape, because many important object properties consist in how the different portions of the shape relate to one another (e.g., whether the object is top-heavy; whether it is symmetrical; whether it is composed of multiple distinct parts). Other researchers have investigated some of these more ‘global’ aspects of shape representation. For example, when looking at objects we can estimate their center of mass (Baud Bovy & Soechting, 2001; Cholewiak, Fleming, & Singh, 2013), report certain kinds of symmetrical relations (Sawada & Pizlo, 2008; for a review see Treder, 2010) or identify different functional and meaningful parts (de Winter & Wagemans, 2006; Hoffman & Richards, 1984). For example, the handle of the pot in Fig. 2 is usually perceived as being a different part and serving a different purpose than the bowl. There is no way we could do that without perceptually organizing local information into more global units and assign meaning to them.

Another important line of research is devoted to the format of the shape representation used by the visual system. A widely recognized approach coming from computer vision (Blum, 1973; see also Feldman and Singh (2006), Twarog, Tappen, and Adelson (2012)) is the medial axis transform. The medial axis of a shape or object (see Fig. 2) can be imagined as its underlying ‘skeleton’—similar to the bones of a human body—which captures the local symmetry axis of the constituent parts. All skeletal representations are organized hierarchically, representing object features at different levels of resolution. Bigger branches (or parent branches) thereby resemble more global object features and smaller branches code the fine structure of objects (local features). Kovacs and Julesz (1994) found that Gabor targets were easier to detect when located on the medial axes of objects. More recently, by asking hundreds of subjects to tap once anywhere within a shape, Firestone and Scholl (2014) found that the pattern of tapping points reflected the shape’s medial axis. Such findings suggest that at very least, medial axes are locations that satisfy important geometrical conditions that are important for visual processing, and possibly reflect explicit representation of medial axes by the visual system.

Despite their potential importance, there is very little research on the interpretation of the meaning of parts or part structure (but see Kim and Feldman (2009), Spröte and Fleming (2013)). In other words, what causal origins do we assign to certain parts of an object and how do these inferred causes influence our perception of shape?
The idea that transformations and causality could be important aspects in shape perception has been the subject of several studies. Inspired by Thompson’s (1992, originally published 1917) ideas about how biological shapes relate to one another through topological transformations, Shaw and Pittenger (1977) and Cutting (1978) showed that the visual system is familiar with the spatial transformations that face profiles undergo during aging. Interestingly, applying similar transformations to profiles of non-face objects (Mark, Todd, & Shaw, 1981; Pittenger, Shaw, & Mark, 1979) can cause them to appear to take on different subjective ‘ages’, suggesting that it is the transformation per se, rather than face-specific features that the visual system has internalized.

The perception of causality has been thoroughly studied in dynamic scenes involving interactions between objects. For example, Michotte (1963) developed hundreds of demos such as the famous ‘launching effect’, in which one circle appears to cause another stationary circle to start moving by colliding with it. Bae and Flombaum (2011) even demonstrated that such a percept of causality can occur behind an occluder, without the actual event being directly observable. In this case however, another directly observable but spatially distinct launching event had to be present in order to bias subjects’ preference of perceived causality for the occluded case. This phenomenon is known as causal capture (Scholl & Nakayama, 2002). Despite this nice demonstration of causality perception in the absence of a local event, this kind of causality is still mainly related to explicitly seen interactions.

We, on the other hand, are interested in the perception of forces and interactions between objects that happened in the past and are not directly visible in the moment of observation.

Leyton (1989, 1992) was one of the first to provide a systematic geometrical framework for interpreting shape features in terms of causal attribution. He showed that all curvature extrema along a shape’s outline have local symmetry axes leading to them. He argues that these local symmetry axis mark the directions along which forces operated to create the observed shape. Therefore he termed this symmetry axis “process inferring symmetry axis” (PISA). Put differently, according to Leyton the presence of an extremum tells us that a force acted on the shape and its identity lets us infer the direction along which the force acted on the shape, namely along its PISA. Were the visual system to follow this line of reasoning to its logical conclusion, it would reduce all shapes to extremum, with a causal history detailing how each protuberance or deviation was created (but see Hendrickx and Wagemans (1999) for a critical review of Leyton’s claims).

However, because of its focus on local curvature extrema, the PISA approach cannot capture global transformations of shape such as the twist that we see in Fig. 3A. Despite local differences between the helical ridges distributed across the surface, the twist has left these signatures in the shape, which we cannot help but interpret as originating from a common cause. In contrast to other kinds of inferences based on inductive thinking the transformation here seems to be an inherent part of our immediate percept.

Fig. 3B shows a set of images in which the process or transformation that gave the object its shape is also pretty salient. We can readily identify which aspects of the shape were caused by the application of some kind of transformation to an underlying shape. It is almost as if the visual system separates the observed shape into distinct causal contributions—the true or underlying shape and the transformation subsequently applied to it (see Fig. 4).

There are other domains in perception where it is thought that we can separate images into distinct causal components. Barrow and Tenenbaum (1977) for example suggested that in lightness perception retinal intensity values are decomposed into separate contributions from illumination and reflectance—a ‘blind source separation’ in the intensity gradient domain.

Therefore it is interesting to ask to what extent subjects can separate shapes into distinct causal contributions that have combined to yield the observed shape. In other words: can observers separate shapes into intrinsic object features and extrinsic transformations? We sought to tackle this question experimentally. Our reasoning was as follows: if subjects can abstract shape transformations from objects, then (1) they should be able to identify a given transformation when it is applied to different objects and (2) they should be able to identify a given object when it has been subjected to different transformations. To test this, we developed two complementary tasks, using ‘bending’ as our example transformation.

First we asked whether subjects can match the degree of bending applied to one object (the ‘Test’), by adjusting the amount of bend applied to a different object (the ‘Match’). If subjects can identify and estimate the degree of bend, they should be able to match it. It is important to note that in what follows, we define the degree (and axis) of bend solely in terms of the generative procedure that we used to create the objects (see Methods), i.e., the ‘programmed’ or ‘intended’ degree of bend introduced by the transformation that we applied. Thus, when we refer to the ‘true’ or ‘veridical’ degree of bend, we are making the assumption that only transformations along the axis defined by us are considered bends.

In a second experiment, we ran an identification task in which subjects were shown a bent version of an object and had to identify to which of several non-bent objects it corresponded. This is similar to many face and object recognition experiments in which observers must determine object identity across some kind of transformation, such as viewpoint, or illumination (Braje,
However, to our knowledge, this kind of task has never been investigated for non-rigid transformations applied to arbitrary 3D objects.

2. Experiment 1 (3D asymmetric matching task)

Experiment 1 tested whether subjects can perceive shape transformations and apply them to different shapes. Therefore we asked subjects to adjust the degree of bend of a standard object (match object) until it appeared to be as bent as each test object.

2.1. Methods

2.1.1. Subjects

17 subjects (4 male and 13 female) between 19 and 33 years (mean age: 24 years and 4 month) participated in this experiment. Subjects were recruited through the university's internal mailing list. All participants reported having normal or corrected to normal vision and were naïve to the purpose of the task. Participation was compensated with 8 €/h or course credit. No other reward was granted. All participants gave written informed consent. Experiments were conducted in accordance with the Declaration of Helsinki and approved by the local ethics committee.

2.1.2. Stimuli

The stimuli consisted of stereoscopic renderings of scenes containing a shape to which different degrees of a bending transformation had been applied. Every scene consisted of a white background with a textured, slightly reflective ground plane and a 3D object floating in the middle of the scene above the plane. Eleven different test objects were created, each of which was subjected to five levels of the bending transformation in equal steps from 0° to 160° (see Fig. 5 for examples). One additional object
was created as a match object, which was subjected to 91 levels of the bending transformation from \(-180^\circ\) to \(180^\circ\) in steps of \(4^\circ\). This object was also used as a twelfth test object. The objects were created in Blender 2.67b according to the following procedure, which started with a single cube. One of the six faces of the cube was randomly selected and a duplicate of the cube was added to the selected face, thereby extending the object along that direction. One of the five empty faces of the duplicate cube was randomly selected and a second duplicate was added to that face. This procedure was repeated until the shape consisted of ten cubes in total. The resulting object was smoothed and then perturbed using a 'Cloud' texture displacement modifier to increase surface variability even further. Finally we added visual 'Dalmation' texture markings to the surface. To guarantee a wide range of variability between objects, the base objects (i.e., before the bend was applied) were individually picked by the authors from a larger set of 100 objects, created this way.

Objects were bent in Blender using ‘Armatures’. As shown in Fig. 6 armatures consisted of two sets of bones, connected with joints that are strung together head to base, like a chain of arrows pointing in the same direction. The first bones of the two sets were in turn attached to one another at their bases thereby pointing in opposite directions. The armature was aligned and scaled (all bones being equally long) to the object’s axis of transformation. This axis, along which the bending was applied, was individually chosen by the authors to maximize its plausibility and was typically approximately aligned with the principal axis of the object. A ‘bend’ in our case can be understood as a transformation that modifies the shape of the object so that its (initially straight) transformation axis becomes curved. More specifically, our ‘bends’ also met three other criteria: (1) curvature along the path was of constant sign, (2) the path was mirror symmetric and (3) the turning angle between the tangents of the end-points of the path was less than \(180^\circ\). The bend was created by changing the angle between the armature bones; Blender automatically modifies the shape of the surrounding mesh accordingly, thereby bending the shape. This is why we express the bends in terms of angles. We considered the degree of bending simply as \(180^\circ\) minus the turning angle between the extensions of the two outermost bones of the armature. Using two bones (one for each of the two directions of the transformation axis starting at its center) would crease the object only at its center, leaving the rest of the axis untouched, thereby producing unnatural folds. Therefore, to distribute the bending over a wider range along the axis and producing a smoother bending, we construct the armature from four bones.

Stereoscopic images were created at a resolution of \(2560 \times 1024\) pixels using the ‘Stereoscopic Camera’ Python Add-on for Blender (Version 1.6.8 http://www.noeol.de/s3d/) by Schneider (November 19th 2014). Objects within a scene were slightly shifted to the left (match objects) or right (test objects) from the camera’s axis of sight to compensate for the deviation of the actual image location from the center of the screen in the

![Fig. 5. Two example test shapes (top and bottom row) with increasing degrees of bend (0°, 40°, 80°, 120° and 160° from left to right). Middle row shows samples of the match shape from 0° to 180° in 40° steps.](image)

![Fig. 6. Process of bending meshes using ‘Armatures’.](image)
renderings (320 pixels or approximately 8.4°). We used the ‘Off-Axis’ camera configuration with a separation of 6.5 blender units and Zero Parallax set to 55.5 blender units, which corresponds to an interpupillary distance (IPD) of 6.5 cm and a viewing distance of 55.5 cm respectively. The objects subtended roughly 4.5–6.5° of visual angle. This matched the physical parameters of our Two-Monitor Wheatstone Stereoscope such that the accommodation approximately matched the vergence angle when fixating the objects.

2.1.3. Apparatus

Stimuli were presented on two 19-in. Dell UltraSharp 1907FP LCD monitors at a resolution of 1280 × 1024 each. Stimuli for the left eye were presented on the left monitor and stimuli for the right eye were presented on the right monitor. To bring the two views into alignment we used a Wheatstone mirror stereoscope. Viewing distance was kept constant at 55.5 cm using a chin rest.

2.1.4. Procedure

On each trial the observer saw a test object on the left and a match object on the right of the stereo display (Fig. 7). Subjects had to adjust the degree of bend applied to the match object until it appeared to be “as bent as” the test object. Bending was adjusted using the arrow cursors in steps of 4°, with the ‘Up-Arrow’ increasing the degree of bend and the ‘Down-Arrow’ decreasing the degree of bend. Pressing the ‘Enter’ key confirmed the adjusted degree of bend and started the next trial. Subjects were informed that the minimum and maximum degree of bend of the match object did not coincide with the minimum and maximum degree of bend in the test objects (i.e., the range of possible adjustments exceeded what was actually required for veridical performance with regard to our stimulus generation method). Test objects were presented in random order and in random orientations within their local x–z-plane in five blocks of 60 trials each (300 trials in total). The azimuth of the test object’s viewpoint was randomly chosen between 60° and 120° from the transformation axis. The elevation stayed constant at 0°. Match objects were always shown from the same viewpoint with an azimuth of 60° and an elevation of 10° from the transformation axis. The initial degree of bend of the match shape on each trial was randomly set to one of the 91 levels of bend between −180° and 180°. Observers were informed about the anonymous treatment of their data as well as their right to abort the experiment at any time without consequences on their reward. After reading the instructions subjects started the first trial by pressing ‘Enter’.

2.2. Results and discussion Experiment 1 (3D asymmetric matching task)

Fig. 8 plots subjects reported match bend angles as a function of the ‘true’ bend angles of the test objects, according to the generative procedure described in the method section, averaged across all observers and objects. We performed linear mixed-effects analysis with ‘true’ degrees of bend as the fixed effect. We had intercepts for ‘shape’ and ‘subject’ as random effects, as well as by-shape and by-subject random slopes for the effect of ‘true’ degrees of bend. The data clearly show that increasing test bends are matched by increasing match bends (β = .41 ± .067° (standard errors), χ²(1) = 17.926, p < .001), although the slope of the function is lower than ideal (perfect) performance (red line).

Adjustment data for the individual shapes (Fig. 9) reveal that this global pattern also holds on the individual shape level for all but one shape (shape #4). A similar linear mixed-effects analysis but for the individual shapes show increases in subjects’ matches between 0.15 ± .054 (standard error; χ²(1) = 6.217, p = .013) and 0.77 ± .025° (standard error; χ²(1) = 68.709, p < .001) for an increase in the ‘true’ degree of bend by one degree. Simply put, for all but one object, subjects bent the match object more, the more the test object was bent.

We also find that there is a systematic relationship between the elongation of the object and the slope of the matching function. The slopes of the individual objects’ functions increased with an

---

Please cite this article in press as: Spröte, P., & Fleming, R. W. Bent out of shape: The visual inference of non-rigid shape transformations applied to objects. Vision Research (2015), http://dx.doi.org/10.1016/j.visres.2015.08.009
increase in shape elongation (measured as the length of the transformation axis; $r = .74$, $R^2 = .55$). In other words, the longer the test shape was, the better subjects were at matching the degree of bend. This makes intuitive sense, since the effects of bends on elongated objects are simpler and more salient than on more compact objects. When applied to elongated objects, even relatively small bends produce clearly visible curvature along the transformation axis. By contrast, for more compact objects, the transformation axis only starts appearing clearly bowed at large values of bend. This relationship between elongation and matching accuracy hints that subjects probably used the extent to which the shape had a bowed, arc-like global structure as one of the main cues to the degree of bend.

To test this idea, we developed a simple model—based on measurements derived from the 3D object meshes—to test how well the overall arc-like shape of each object could predict subjects’ performance. The arc shape results from the specific configuration of armatures used to create the bends in each object, which were placed along that object’s global/principle axis. These axes are similar (though not strictly identical) to the first linear principal components of the objects’ 3D meshes (see Methods section for details). However, since linear PCA by definition cannot capture non-linear patterns, we used non-linear principal component analysis (version 0.88 of the Nonlinear PCA toolbox for MATLAB by Scholz (2012)) to estimate the axes of the objects from their full 3D meshes after the bends were applied. We then determined the match shapes whose curvature histograms (i.e. distribution of curvature values along the axis) were most similar (smallest Earth Mover Distance) to those of the test shapes. This allowed us to derive a predicted match for each of the test objects (orange lines in Fig. 9).

The differences in the predictions of our model (NLPCA) and ground truth show that the axes derived by NLPCA and the ‘true’ axes defined by the armatures are not necessarily identical. This shows that estimating an objects’ curved global axis is not trivial, even given a complete 3D surface description. NLPCA is affected by all portions of the mesh equally, even those that are not strongly affected by the bending process, which can cause deviations in the estimated axis shape. In contrast, the visual system may be able to up- or down-weight the contribution of different object parts to its global structure. Fig. 10 however shows that the NLPCA model nevertheless captures at least as much about the pattern of errors made by subjects as ground truth. More importantly, except for maximally two cases (shape# 2 and 3, $t_{(16)} > 5.30$, $p < .001$), the NLPCA based model makes the same amount (shape# 5, 6, 7, 10 and 12, $t_{(16)} < 1.9$, $p > .075$) or even less errors (shape# 1, 4, 8, 9 and 11, $t_{(16)} > 2.67$, $p < .017$) in predicting subjects’ performance than ‘ground truth’. In addition, while ‘ground truth’ cannot by

---

Please cite this article in press as: Spröte, P., & Fleming, R. W. Bent out of shape: The visual inference of non-rigid shape transformations applied to objects. Vision Research (2015), http://dx.doi.org/10.1016/j.visres.2015.08.009
bend in Figs. 8 and 9, then we should be able to measure this effect was applied, it could be that even the unbent objects contain axial degree of bend in the match object. Even though no explicit bend direction, and therefore need to compensate by over-stating the effect is specific to the stimulus being viewed. The set of stimuli was identical to the match stimuli from the previous experiment except that now all 12 base objects were available in 91 levels of bend from $-180^\circ$ to $180^\circ$ in steps of $4^\circ$.

3.1. Subjects
The same group of subjects participated as in the previous task. Participation was again compensated with 8 €/h or course credit. No other reward was granted. All participants gave written informed consent. Experiments were conducted in accordance with the Declaration of Helsinki and approved by the local ethics committee.

3.1.2. Stimuli
On each trial an observer saw a single 3D object in the center of the stereo display. Subjects had to adjust the degree of bend applied to the object until it appeared to them as being neutral (i.e., not bent). Bending was again adjusted using the arrow cursors in steps of $4^\circ$, with the ‘Up-Arrow’ increasing the degree of bend and the ‘Down-Arrow’ decreasing the degree of bend. Pressing the ‘Enter’ key confirmed the adjusted degree of bend and started the next trial. Objects were presented in random order in five blocks of 12 trials each (60 trials in total). Objects were always shown from the same viewpoint with an azimuth of $60^\circ$ and an elevation of $10^\circ$ from the transformation axis. The initial degree of bend of the match shape was randomized. Observers were informed about the anonymous treatment of their data as well as their right to abort the experiment at any time without consequences on their reward. After reading the instructions subjects started the first trial by pressing ‘Enter’.

3.1.3. Procedure
Subjects perceived the objects to be unbent when they were actually bent by $39.58^\circ$ on average. This suggests that when displayed in their neutral state (according to the generative process), subjects perceived the objects as being bent to some extent in one direction, which could be corrected by applying a bend in the other direction. However, except for one case (shape 11), all blue bars are higher than the orange bars. This means that subjects’ adjustments in response to neutral (not bent) objects in the ‘asymmetric matching task’ resulted in higher degrees of bend than adjustments in the ‘nulling task’. This is supported by a paired sample $t$-test between the adjustments of the two experiments across all shapes ($t(11) = 3.850, p = .003$). Paired sample $t$-tests for the individual shapes reveal significant differences for seven out of the twelve comparisons ($t(16) \geq 2.692, p \leq .016$) with another two comparisons approaching significance (object 3: $t(16) = 1.927, p = .072$ and object 10: $t(16) = 2.051, p = .057$). Moreover, we did not find a significant correlation between the errors in the ‘asymmetric matching task’ and the ‘nulling task’. In most cases, subjects are not consistent in their errors across the different objects (except subjects 5 and 6, $r \geq .62, p \leq .03$) nor do the errors for the individual objects across subjects correlate with each other (except object

![Fig. 10. Comparison of errors of our model (NLPCA) and ground truth (i.e., perfect performance) in predicting subjects’ data for each shape (dots with numbers indicating the shape number) expressed as the ratio between the errors of the NLPCA model and ground truth. White dots indicate shapes for which the errors for the two models differed significantly (i.e., where one model was better at predicting subject’s data than the other; $p < .05$); black dots indicate where there is no significant difference in errors.](Image 38x486 to 278x726)

Fig. 10. Comparison of errors of our model (NLPCA) and ground truth (i.e., perfect performance) in predicting subjects’ data for each shape (dots with numbers indicating the shape number) expressed as the ratio between the errors of the NLPCA model and ground truth. White dots indicate shapes for which the errors for the two models differed significantly (i.e., where one model was better at predicting subject’s data than the other; $p < .05$); black dots indicate where there is no significant difference in errors.

... definition capture any variation in the slope and intercept of subjects’ performance, the NLPCA based model does (52% explained variance for the slope and 41% for the intercept).

Taken together, our modeling suggests that inferring the degree of bend from images of bent objects is not trivial. NLPCA not only captures deviations from straightness due to bending but also due to other inherent object features (e.g. additional limbs). Deviations of this kind could be a potential source for errors also made by our subjects. However, there are also a number of other possible explanations why subjects’ responses understate the effects of changes in the degree of bend. Subjects overestimate the degree of bend when the transformation was low and often slightly underestimate it when the transformation was high. One possibility could be a stimulus independent response bias that is often observed with subjective ratings. Specifically, subjects may be somewhat reluctant to use the ends of the scale and therefore the data tends to regress to the mean (central tendency). However, this explanation seems unlikely when we take into account how subjects performed when the test object was identical to the match object (object 12), apart from a random orientation difference. Here, subjects use almost the whole range of the scale showing no sign of a bias, suggesting that the effect is specific to the stimulus being viewed.

A second potential explanation is that subjects actually perceive the unbent objects as being bent to a certain degree in the opposite direction, and therefore need to compensate by over-stating the degree of bend in the match object. Even though no explicit bend was applied, it could be that even the unbent objects contain axial curvature that is interpreted as a certain amount of bend. If this is the origin of the overestimation we observed for small values of bend in Figs. 8 and 9, then we should be able to measure this effect by asking subjects to adjust each test object until they appeared to be neutral (i.e., not bent). Experiment 2 tested this possibility using a nulling task in which we asked subjects to adjust the degree of bend of the objects from Experiment 1 until they appeared to not to be bent.

3.2. Results and discussion Experiment 2 (nulling task)

Fig. 11 shows how bent the individual objects were at the point when subjects perceived them to be neutral (not bent), contrasted with the corresponding adjustments in the asymmetric matching task.

Subjects perceived the objects to be unbent when they were actually bent by $39.58^\circ$ on average. This suggests that when displayed in their neutral state (according to the generative process), subjects perceived the objects as being bent to some extent in one direction, which could be corrected by applying a bend in the other direction. However, except for one case (shape 11), all blue bars are higher than the orange bars. This means that subjects’ adjustments in response to neutral (not bent) objects in the ‘asymmetric matching task’ resulted in higher degrees of bend than adjustments in the ‘nulling task’. This is supported by a paired sample $t$-test between the adjustments of the two experiments across all shapes ($t(11) = 3.850, p = .003$). Paired sample $t$-tests for the individual shapes reveal significant differences for seven out of the twelve comparisons ($t(16) \geq 2.692, p \leq .016$) with another two comparisons approaching significance (object 3: $t(16) = 1.927, p = .072$ and object 10: $t(16) = 2.051, p = .057$). Moreover, we did not find a significant correlation between the errors in the ‘asymmetric matching task’ and the ‘nulling task’. In most cases, subjects are not consistent in their errors across the different objects (except subjects 5 and 6, $r \geq .62, p \leq .03$) nor do the errors for the individual objects across subjects correlate with each other (except object
10, r = .71, p = .04) in the two tasks. In other words, an overestimation in Experiment 1 did not lead to a similar amount of compensation in the opposite direction in Experiment 2. This suggests, that the consistent over-estimation for low values of bend that we observed in the asymmetric matching task is not because subjects see the unbent objects as bent—at least not along the direction that the bend was actually applied. However it should be noted that there are some potentially important qualitative differences between the two methods. In Experiment 1 subjects matched static test objects with a dynamic match object, whereas in Experiment 2, they adjusted the neutral point of the same test object dynamically. Although this may sound like a subtle distinction the presence of continuous changes in shape may have provided crucial additional information (but see Kourtzi and Shiffrar (2001)). For the dynamic versions of the test objects in Experiment 2, kinematic restrictions (bending in a specific direction) unambiguously define the direction of the bend transformation. By contrast in Experiment 1, where the test objects were static, no such cues are available, so the direction of the transformation was unknown. Therefore subjects may have seen the low-bend objects in Experiment 1 as being bent along some other direction than the true direction of the transformation. It is also worth noting that using the unsigned mean causes random variance to accumulate, which may overstate the extent of the errors in the nulling task. However, at least in some cases, the large deviations (e.g. over 45°) probably indicate that the errors truly reflect a bias rather than noise.

Taken together the findings of Experiments 1 and 2 suggest that subjects can to a certain extent match bends applied to different objects. When the bends that we applied were only weak, and therefore not vary salient, the subjects tended to interpret the intrinsic curvatures of the object as being caused by bends, even when they were not introduced by the specific generative process that we applied. We believe the over-estimation of bends for the weakly-bent stimuli was due to subjects interpreting the objects as being bent to a certain degree but in a different way (i.e., along a different direction) to the one we applied. This resulted in flatter matching functions in Experiment 1. Nevertheless, for many shapes, once the transformation was large and salient, subjects readily perceived and matched it.

If our initial suggestion, that the visual system can separate out underlying generative processes is true, than subjects should also be able to discount information about the transformation when identifying objects. We know that they can do this to some extent in other contexts. For example, face and object recognition show a certain degree of invariance across various transformations (Braje et al., 1998; Tarr & Pinker, 1989). In Experiment 3 subjects were shown test objects in random orientations and were asked to identify the corresponding object from a set of four candidates in different orientations. Importantly, while the candidate objects were always non-bent versions of the objects, in half the trials, the test object was bent (see Fig. 12). Thus, to identify the matching object, subjects had to be able to visualize not only the effects of rigid rotations, as in standard mental rotations experiments (Shepard & Metzler, 1971), but also the effects of non-rigid transformation. Failing that, to succeed at the task, they would have to be able to rule out the false candidates by determining that they would not match the test object no matter how they were bent.

4. Experiment 3 (4AFC identification task – target present)

4.1. Methods

4.1.1. Subjects

A new group of 15 subjects (3 male and 12 female) between 20 and 30 years (mean age: 24 years and 1 month) participated in this experiment. Subjects were recruited through the university’s internal mailing list asking for participation in a 3D shape perception task. All participants reported having normal or corrected to normal vision and were naive to the purpose of the task. Participation was compensated with 8 €/h or course credit. No other reward was granted. All participants gave written informed consent. Experiments were in accordance with the Declaration of Helsinki and approved by the local ethics committee.

4.1.2. Stimuli

The set of stimuli consisted of 99 3D scenes (88 renderings of the test objects and 11 renderings of the same base objects but with different 3D orientations), similar to the ones used in the previous experiments. Scenes again consisted of a textured, slightly specular ground plane and a 3D object floating above the plane. The set of test objects was comprised of the same 11 base objects used in Experiments 1 and 2, with two bending levels (0° and 160°) and four different orientations. One orientation was chosen randomly and the other three in ascending steps of 90° around a 360° circle. The four orientations were the same across the different test objects. Test objects created from the same base shape were viewed from an angle with an azimuth of 60°, 80°, 100° or 120° randomly chosen but balanced across the 8 configurations per object (two levels of bend × four orientations) and an elevation of 10° from their transformation axis. Identification objects were identical to the 11 base objects viewed from an angle with an azimuth of 60° and an elevation of 10° from a transformation axis determined for the same base object among the test objects.

4.1.3. Procedure

Stimuli were presented using the same stereo setup as in the previous experiment. On each trial an observer saw a test object in the middle of the active screen surrounded by four identification objects (see Fig. 12). Identification objects (‘candidates’) were
presented in the four corners of the active screen, 16.86° from the center. Subjects had to indicate which candidate object was identical to the test object in terms of shape, while ignoring differences in orientation and degree of bend. By pressing one of four buttons the object location assigned to that button was surrounded by a green frame. Subjects confirmed their response and started the next trial by pressing ‘Enter’. Test images were presented in random order in four blocks of 88 trials (352 in total). The target object (identification object with the same shape as the test object) was presented randomly at one of the four locations surrounding the test object balanced within a block. The remaining three locations were filled with distractors (different shape than the test object) randomly selected from the ten remaining identification objects.

4.2. Results and discussion

In Experiment 1 we found that subjects can identify transformations across objects (at least to some extent). With Experiments 3 we sought to test whether subjects can also identify objects across different transformations. In other words, can they identify the correct base object (unbent version of an object) from a set of four base objects, whose potential manipulation would lead to the appearance of the test object.

Fig. 13 shows the mean proportion correct for each subject across shapes, separated by whether the test object was bent (orange bars) or not (blue bars). One sample t-tests confirmed that all subjects performed significantly (t(10) ≥ 4.638, p ≤ .001) above chance (25% correct answers indicated by red dashed line). This suggests that subjects can discount bending transformations applied to unfamiliar objects.

Additional paired sample t-tests showed that nine out of 14 subjects identified the correct base object significantly (t(10) ≥ 2.466, p ≤ .034) more often when the test objects were unbent as opposed to bent (see Fig. 13, asterisks indicate level of significance). The result across all subjects reflects the pattern already observed for individual subjects. The mean proportion of correct identifications across subjects was 0.77 (SEM = 0.02) for unbent and 0.59 (SEM = 0.03) for bent objects. An additional paired sample t-test confirmed that subjects are on average significantly better in identifying the correct base shape when the test shape was unbent in contrast to bent (t(13) = 8.669, p < .001), but both were significantly above chance. Put simpler, unbent test objects were easier to identify than bent ones. This indicates, unsurprisingly, that there is an additional cost to undoing the non-rigid transformation compared to simply mentally rotating the object.

To see whether errors in subjects’ responses were systematic across objects, we looked at confusions between the correct (targets) and incorrect base objects (distractors). Fig. 14 illustrates this frequency of target confusions with distractors weighted by the frequency of target-distractor co-occurrences. Put differently, how often did subjects ‘confuse’ the correct base object (target) with a specific incorrect base object (distractor)? As already evident in the mean results shown in Fig. 13 the diagonal line in the left and right panel in Fig. 14 exhibit high values reflecting high rates of correct identifications. When confronted with unbent test objects (left panel) subjects confused object six with object three and object seven with object six significantly above chance (25%). For bent test objects, objects three, four, eight and nine were confused significantly above chance with objects ten, nine, one and again one respectively. In short, we found that certain distractors were significantly often confused with certain targets. Fig. 15 shows that the objects subjects most often confused with each other also perceptually appear similar in terms of their shape. In other words, subjects made most of their errors when objects were perceptually similar.

However, this raises the question about the strategy subjects used in order to do the task. It is possible that some of the successes in this task might have been achieved by a process of elimination, rather than by positively identifying the correct target. At least in some situations (low shape similarity), rather than picking an object because subjects knew it was the correct one, they picked an object because the alternatives were just less likely than the correct ones. The higher the perceived shape similarity, the less
reliably they would be able to rely on that strategy. Experiment 4 sought to reduce this potential to correctly identify an object in Experiment 3 by mainly relying on evidence against the other objects (‘exclusion strategy’) rather than deliberately choosing this object based on positive evidence for it. In other words, when subjects in Experiment 3 would have relied on ‘exclusion’ to pick a target when no object accumulated sufficient positive evidence (to be chosen as the target), subjects in Experiment 4 were given the option not to pick a target at all (and less often rely on ‘exclusion’).

5. Experiment 4 (4AFC identification task – target absent)

5.1. Methods

5.1.1. Subjects

A new group of 17 subjects (7 male and 10 female) between 20 and 37 years (mean age: 24 years) participated in this experiment. Subjects were recruited through the university’s internal mailing list asking for participation in a 3D shape perception task. All participants reported having normal or corrected to normal vision.

Fig. 13. Proportion correct identifications for all 14 subjects as well as their mean for unbent and bent shapes. The red dashed line indicates chance performance. Error bars indicate SEM across the 11 different shapes. Subjects are well above chance in identifying the correct target among distractors, with identification performance being significantly higher for unbent vs. bent shapes. Stars indicate the level of significance: no star, $p > .05$; 1 star, $p < .05$; 2 stars, $p < .01$; 3 stars, $p < .001$.

Fig. 14. Weighted confusions of targets with distractors during identification. Absolute values of confusion were weighted by the frequency of target-distractor co-occurrences. The left panel shows the confusion matrix for unbent, the right panel the matrix for bent test objects. Red squares indicate confusions significantly different from chance (25%). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Please cite this article in press as: Spröte, P., & Fleming, R.W. Bent out of shape: The visual inference of non-rigid shape transformations applied to objects. Vision Research (2015), http://dx.doi.org/10.1016/j.visres.2015.08.009
Fig. 15. Distractors from the identification task that were confused with the target objects for the unbent (A) and the bent (B) condition. The four horizontally aligned objects under 'test object #' represent different views of the same object.
and were naive to the purpose of the task. Participation was compensated with course credit. No other reward was granted. All participants gave written informed consent. Experiments were in accordance with the Declaration of Helsinki and approved by the local ethics committee.

5.1.2. Stimuli

We used the same stimuli as in the previous experiment.

5.1.3. Procedure

The procedure was the same as in the previous experiment (see Fig. 12) except that in half of the trials no target was presented, only distractors. Subjects were informed about the frequency of target-absent trials. By pressing ‘Spacebar’ observers indicated target-absent trials and the test object was in turn highlighted in green.

5.2. Results and discussion

Since in only 50% of the cases the correct target was present and subjects either responded by picking an object or rejecting the entire set of candidates, we first analyzed performance in terms of ‘target present’ versus ‘target absent’ classifications. Note that a ‘target present’ classification is not synonymous with identifying the correct target. It just means that subjects decided to pick any object (even an incorrect one) whenever the target was present in the display. One sample t-tests with a test value of 0.5 (chance performance) confirmed that all subjects performed significantly above chance (t(10) ≥ 2.246, p ≤ .049). Additionally, paired sample t-test indicated that 11 out of 17 subjects were significantly (t(10) ≥ 2.38, p ≤ .039) better in classifying trials where the test object was unbent as opposed to bent (see Fig. 16), as in the previous experiment.

The mean proportion of correct classifications across subjects was .76 (SEM = .02) for trials with unbent test objects and .58 (SEM = .01) for trials with bent test objects. A paired sample t-test confirmed that subjects are on average significantly (t(16) = 7.075, p < .001) better in classifying trials with unbent test objects than trials with bent test objects.

Fig. 17 plots the difference between errors (Misses–False alarms) made by subjects in unbent test object and bent test object trials. Visual inspection already suggests that there seems to be a fundamental difference in the type of errors responsible for incorrect responses between the two types of transformations. A paired sample t-test and two one-sample t-tests confirm that the mean difference between Misses and False Alarms across subjects is significantly higher (t(10) = −8.077, p < .001) for bent test objects trials (orange bar, mean difference = .35) than for unbent test objects (blue bar, mean difference = −.06) and that only the former significantly differs (t(10) = 4.288, p = .001) from a test value of 0 (identical amount of Misses and False Alarms). This means that the frequency of the two types of errors is balanced for unbent test object trials, whereas errors in bent test object trials are dominated by Misses. Put differently, only when the test object is bent subjects commit one type of error more often, namely Misses – they fail to identify that the target is present, suggesting they do not successfully identify the object across transformations.

Whenever they decided to select an object as the target when the target was actually present (see Fig. 18), subjects on average performed significantly (two one sample t-test with t(10) ≥ 9.178, p < .001, test value = .25) and considerably above chance. Again, unsurprisingly, identification performance in unbent test object trials significantly exceeds bent test object identification (paired sample t(16) = 6.258, p < .001), similar to Experiment 3.

Taken together the results shown in Figs. 16 and 17 suggest that at least for unbent test objects the task of identifying the correct base object was much more difficult than would be expected from the results of Experiment 3 (see Fig. 13). Subjects perform at a relatively low level overall because they often reject target present displays (Misses) under uncertainty. However, when the subjects did spot a target, they were usually correct, suggesting that for at least some shapes the effects of the transformation were easy for the subjects to predict.

Fig. 16. Proportion correct responses whether the target is present in the display for all 17 subjects for unbent and bent shapes. Error bars and red dashed line indicate SEM and chance performance respectively. Asterisks indicate significant differences between unbent and bent objects. Note that due to the compound nature of this paradigm a correct ‘detection’ only requires any identification response – subjects pick any object as opposed to a target absence response. This also includes false identifications. Subjects classify the presence of target objects significantly above chance. They also classify unbent object trials significantly better than bent object trials. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
6. General discussion

In a previous study (Spröte & Fleming, 2013), we found that subjects can interpret concavities in objects as being caused by ‘bites’ (i.e., the forceful excision of material from an object). Here we investigated the perception of other processes that can introduce concavities, namely ‘bends’ (i.e., non-rigid deformation of objects). In Experiment 1, we found that subjects can to a certain extent extract information about bending transformations applied to shapes while ignoring other differences. They could approximately estimate and transfer the magnitude of the bend from one base shape to another. This is a non-trivial ability as it requires attributing some of the properties of the test object to a transformation that is combined to an unknown extent with the original features of the shape. The subjects had a tendency to overestimate the amount of bend for the least bent objects (and to a lesser extent to underestimate the amount of bend for larger values). A small proportion of this effect may be due to a response bias caused by a tendency to avoid the ends of the scale. However, it is also possible that the bias was caused by interpreting the low-bend objects as being bent to some extent. The results of the nulling task (Experiment 2) suggest that in most cases, the residual perceived bend could not predict the over-estimation of small bends observed in Experiment 1. This suggests that if subjects do interpret some of the curvatures of the low-bend objects as being due to a bending transformation, it is a transformation along some other axis than...
At this point we can only speculate about potential factors that determine the level of abstraction and limit the completeness of these perceptual and cognitive inferences. Intuitively, the amount of experience with the specific type of transformation presumably plays an important role. Another potential limiting factor could be interactions between the complexity of the initial shape and the transformation. Some transformations have no effect on any shape (e.g., rigid rotation) while some have effects on most shapes, but not on specific shapes (e.g., twisting a sphere around an axis that passes through its center will result in a sphere). In these cases, inferring transformations from shape alone is obviously impossible as they leave no detectable trace in the shape. However, even when the transformations do have measurable effects on a shape, there are nevertheless quite complex interactions between initial shape and transformation, which limit the extent to which we can separate them. For example, we might expect transformations that preserve part structure and symmetry to be less salient than those that disrupt symmetries or introduce or remove parts (Barenholtz, Cohen, Feldman, & Singh, 2003). Moreover, as either the initial shape or the transformation applied to it becomes very complex, it may become harder to fully invert the causal history. For example, shapes that are initially complex may make it hard to work out which elements have been affected to what extent by a given transformation. This may have influenced subjects’ responses in our experiments to some extent. We find that the visually ‘more complex’ shapes in our experiments (especially shapes 4, 8 and 10) were the ones for which subjects tended to perform worst, especially when the part configurations could have been interpreted as a ‘bend’ rather than as two parts attached to each other. Moreover, applying more complex transformations (e.g., crumpling) or a whole sequence of transformations can also render characteristic features introduced by previous transformations invisible so that recovery is incomplete or impossible. Despite these difficulties there are many cases, including in our experiments, where we readily infer that particular shape features result from specific causal processes.

It remains unclear what strategy subjects relied to perform the task. In other words, how did the subjects achieve transformation-invariance? Some important insights could come from research on view-invariance in object recognition. The main question addressed in such research is how we recognize objects from different perspectives despite dramatic changes in the retinal images. View-invariance can be thought of as a special case of transformation-invariance, in which the transformation is rigid.

One approach to view-invariant object recognition assumes that a template—an internal 3D model of an object—is compared to an external image through a process of mental alignment (Huttenlocher & Ullman, 1987). Shepard and Metzler (1971) showed that comparisons between objects in different orientations tend to take longer, the higher the angular difference between two views of the same object. This, and other results, strongly suggests that participants can perform an explicit mental imagery of the rotation transformation, which is analogous to a real physical rotation. If we extend such an idea to our identification task, this would mean that subjects might go through a process of mental imagery, like in mental rotation, where they ‘bend’ or ‘unbend’ a mental image of the object in their mind. It would be interesting to perform analogues of the classic Shepard and Metzler experiments with non-rigid transformations to test this hypothesis.

A second group of models assumes that the visual system exploits features that stay essentially invariant across changes in viewpoints (Biederman, 2001; Cooper, Biederman, & Hummel, 1992). Translating this concept into the realm of non-rigid shape transformations, there are clearly many aspects of the shape which are essentially unaffected by the bend transformation.
example, subjects could try to find correspondences in the local texture marking, the number of limbs of the object, or other salient features that are conserved across bends, and identify target objects based on these invariants.

As it stands, our study does not favor either of these possible approaches over the other. It may indeed be that subjects use strategies that combine, or lie somewhere between these two alternatives. On the one hand, subjects in the identification task sometimes relied on a process of elimination. This indicates that at least under certain circumstances mental unbending does not lead to an unambiguous result. On the other hand, the same finding weakens a strong argument for the invariance approach. It would suggest that the features subjects rely on are only invariant across a certain range of the transformation and are only invariant in certain objects. This would be in line with the abovementioned limitations of ‘shape scission’, namely that inferences can occur at different levels of resolution, with certain shapes yielding richer and more detailed inferences than others.

Another issue that remains unclear is the extent to which the causal inferences made in our study are perceptual or cognitive in nature. Certainly the phenomenology of the objects in Figs. 1 and 3, for example, tend to suggest there is at least some causal inference made in our study are perceptual or cognitive and more detailed inferences than others.

It would be at least under certain circumstances mental unbending does not lead to an unambiguous result. On the other hand, the same finding weakens a strong argument for the invariance approach. It would suggest that the features subjects rely on are only invariant across a certain range of the transformation and are only invariant in certain objects. This would be in line with the abovementioned limitations of ‘shape scission’, namely that inferences can occur at different levels of resolution, with certain shapes yielding richer and more detailed inferences than others.

More generally, we suggest that the inference of transformations applied to objects is one example of a broader class of perceptual-cognitive judgments about objects derived from their shape, which we call ‘Shape Understanding’. We suggest that whenever we view a novel object, the visual system uses perceptual organization processes to infer a primitive ‘generative model’ that captures the object’s key features and parameters its natural degree of variation. Such generative models would be a sort of ‘recipe’ describing the shape’s underlying logic and causal origins, allowing us to understand shape somewhat like parsing a sentence allows us to understand its meaning. Inferring generative models of shapes would potentially enable us to make a number of other potentially useful inferences about an object, derived from its shape. For example, in this article, we have focused on how the visual system infers the causal history, or transformations, that are responsible for giving the shape its most important features. However, such inferences are unlikely to be limited to simple geometrical or physical transformations like bending or twisting, but probably include a wide range of other processes that create and modify shape, such as fluid flow, artificial manufacture, biological growth, aging and weathering, etc. Moreover, in addition to inferring the forces and processes that have acted on the object, we can also make many judgments about the object’s material properties from its shape (e.g. inferring a liquid’s viscosity from the shapes it adopts as it flows in response to gravity, Paulun, Kawabe, Nishida, and Fleming (2015)). This in turn would help us to work out how the object or material would be likely to respond to future interactions. Being able to predict the most plausible ways in which a shape is likely to vary (e.g., elongation of parts) could also allow us to predict what variants of the shape would be likely to look like: i.e., to infer the plausible shape of other members of the same class of object, even when given only a single exemplar (Fleming, 2015). Thus, by combining perceptual organization processes with cognitive inference, ‘Shape Understanding’ may enable us to derive a wide range of important inferences about unfamiliar objects based on their shape.

To conclude, we showed that shape representations include at least some limited information about the transformations or processes that contributed to the observed shapes. This ability to infer generative processes has far reaching benefits for the observer. It may facilitate the prediction of future behavior of the object—how it would likely respond to other forces and processes. Moreover, inferring generative models of shapes may aid the categorization of novel objects from limited numbers of exemplars, by enabling the observer to predict what plausible variations of the objects might look like. We believe this raises many interesting questions for perceptual organization research.

Acknowledgments

The authors wish to thank Steven Cholewiak for invaluable comments throughout the research project. Part of this research was funded as part of the Collaborative Research Centre on ‘Cardinal Mechanisms of Perception’, SFB TRR-135 from the DFG (German research foundation).

References

Dretske, F. (2007). Natural degrees of variation. Such generative models would be a sort of ‘recipe’ describing the shape’s underlying logic and causal origins, allowing us to understand shape somewhat like parsing a sentence allows us to understand its meaning. Inferring generative models of shapes would potentially enable us to make a number of other potentially useful inferences about an object, derived from its shape. For example, in this article, we have focused on how the visual system infers the causal history, or transformations, that are responsible for giving the shape its most important features. However, such inferences are unlikely to be limited to simple geometrical or physical transformations like bending or twisting, but probably include a wide range of other processes that create and modify shape, such as fluid flow, artificial manufacture, biological growth, aging and weathering, etc. Moreover, in addition to inferring the forces and processes that have acted on the object, we can also make many judgments about the object’s material properties from its shape (e.g. inferring a liquid’s viscosity from the shapes it adopts as it flows in response to gravity, Paulun, Kawabe, Nishida, and Fleming (2015)). This in turn would help us to work out how the object or material would be likely to respond to future interactions. Being able to predict the most plausible ways in which a shape is likely to vary (e.g., elongation of parts) could also allow us to predict what variants of the shape would be likely to look like: i.e., to infer the plausible shape of other members of the same class of object, even when given only a single exemplar (Fleming, 2015). Thus, by combining perceptual organization processes with cognitive inference, ‘Shape Understanding’ may enable us to derive a wide range of important inferences about unfamiliar objects based on their shape.

To conclude, we showed that shape representations include at least some limited information about the transformations or processes that contributed to the observed shapes. This ability to infer generative processes has far reaching benefits for the observer. It may facilitate the prediction of future behavior of the object—how it would likely respond to other forces and processes. Moreover, inferring generative models of shapes may aid the categorization of novel objects from limited numbers of exemplars, by enabling the observer to predict what plausible variations of the objects might look like. We believe this raises many interesting questions for perceptual organization research.