Cognitive Psychology 90 (2016) 48–70

Contents lists available at ScienceDirect

Cognitive Psychology

journal homepage: www.elsevier.com/locate/cogpsych

# Visual perception of complex shape-transforming processes

Filipp Schmidt *, Roland W. Fleming

*Justus-Liebig-University Giessen, Germany*

ARTICLE INFO

*Article history:*
Accepted 26 August 2016
Available online 12 September 2016

*Keywords:*
Non-rigid transformations
Perceptual organization
Causal history
Object constancy
Shape perception
Shape understanding

ABSTRACT

Morphogenesis—or the origin of complex natural form—has long fascinated researchers from practically every branch of science. However, we know practically nothing about how we perceive and understand such processes. Here, we measured how observers visually infer shape-transforming processes. Participants viewed pairs of objects ('before' and 'after' a transformation) and identified points that corresponded across the transformation. This allowed us to map out in spatial detail how perceived shape and space were affected by the transformations. Participants' responses were strikingly accurate and mutually consistent for a wide range of non-rigid transformations including complex growth-like processes. A zero-free-parameter model based on matching and interpolating/extrapolating the positions of high-salience contour features predicts the data surprisingly well, suggesting observers infer spatial correspondences relative to key landmarks. Together, our findings reveal the operation of specific perceptual organization processes that make us remarkably adept at identifying correspondences across complex shape-transforming processes by using salient object features. We suggest that these abilities, which allow us to parse and interpret the causally significant features of shapes, are invaluable for many tasks that involve 'making sense' of shape.



* Corresponding author at: Justus-Liebig-University Giessen, General Psychology, Otto-Behaghel-Str. 10F, 35394 Gießen, Germany.

*E-mail addresses:* Filipp.Schmidt@psychol.uni-giessen.de (F. Schmidt), Roland.W.Fleming@psychol.uni-giessen.de (R.W. Fleming).

http://dx.doi.org/10.1016/j.cogpsych.2016.08.002
0010-0285/

## 1. Introduction

Every object in our environment—whether a shoe, a starfish or a dollop of cream—has a specific shape, which we readily perceive and can use for recognition and categorization. But where do those shapes come from in the first place? All physical objects and materials end up with particular shapes due to some kind of *generative process*, such as manufacture, biological growth, or self-organization. These shape-forming, morphogenic processes have long fascinated and beguiled researchers from practically every branch of science. However, despite dramatic advances in our understanding of how physical (Chen, Wen, Janmey, Crocker, & Yodh, 2010; Ferziger & Peric, 2012), chemical (Fahlman, 2011; Inostroza-Brito et al., 2015) and biological (Boettiger, Ermentrout, & Oster, 2009; Carlson, 2013; Paluch & Heisenberg, 2009) processes generate complex shapes, we still know surprisingly little about how we perceive and understand such processes (Atit, Shipley, & Tikoff, 2013; Bedford & Mansson, 2010; Chen & Scholl, 2015; Cutting, 1982; Dubinskiy & Zhu, 2003; Feldman, 1995; Feldman & Singh, 2006; Feldman et al., 2013; Hoffman & Richards, 1984; Koffka, 1935/1965; Leyton, 1989, 1992; Mark, Shaw, & Pittenger, 1988; Mark & Todd, 1985; Ons & Wagemans, 2011; Pittenger & Shaw, 1975; Shaw & Pittenger, 1977).

Understanding and inferring shape-transforming processes presumably involves both perceptual and cognitive abilities. Here, we sought to investigate a specific role that perceptual organization processes play in structuring these inferences. In particular, we suggest that one key component lies in computing how locations on objects shift in space as a result of the transformation. Intuitively, to estimate how points are affected by a transformation, the visual system simply has to identify features on the shape that match up across the transformation (Fig. 1).

However, in practice, identifying such features with arbitrary objects, and then estimating dense correspondence for intervening locations is computationally extremely challenging (Fischer & Modersitzki, 2008; Oliveira & Tavares, 2014). Even computer algorithms that are good at establishing point (location) correspondences across non-rigid transformations (e.g., Ma, Zhao, & Yuille, 2016; Movahedi & Elder, 2010; Myronenko & Song, 2010) eventually fail on images that we find intuitively easy to understand (Fig. 2). Moreover, these algorithms generally involve slow and costly iterative computations, and are fragile, often requiring manual parameter tweaking to achieve optimal results for a given shape and transformation. Furthermore, such algorithms are designed for identifying correspondences between features located on the object itself, but do not usually infer how arbitrary points in space are affected by the transformation. This contrasts with the apparent robustness and flexibility of the human visual system in solving these tasks, which we demonstrate here.

Indeed, the subjective ease with which we tend to solve such problems belies their underlying difficulty, and is only loosely related to transformation complexity. For example, mathematically 'simple' transformations like reflections are not necessarily perceptually simpler to solve (e.g., Gregory & McCloskey, 2010), while at the same time, we seem to be extremely robust at distinguishing causally significant shape features from those that are due to noise, which is a computationally challenging

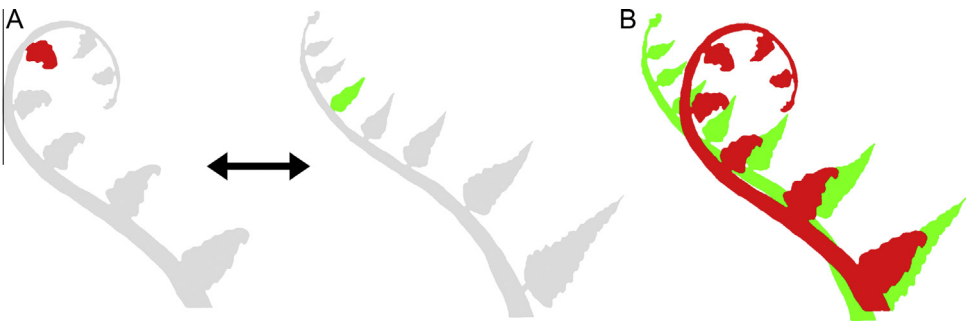


**Fig. 1.** (A) Example of intuitive matching of shape features across non-rigid transformation (i.e., uncoiling of a fern). (B) Superimposed shapes show that this correspondence cannot be established by simple template mapping.

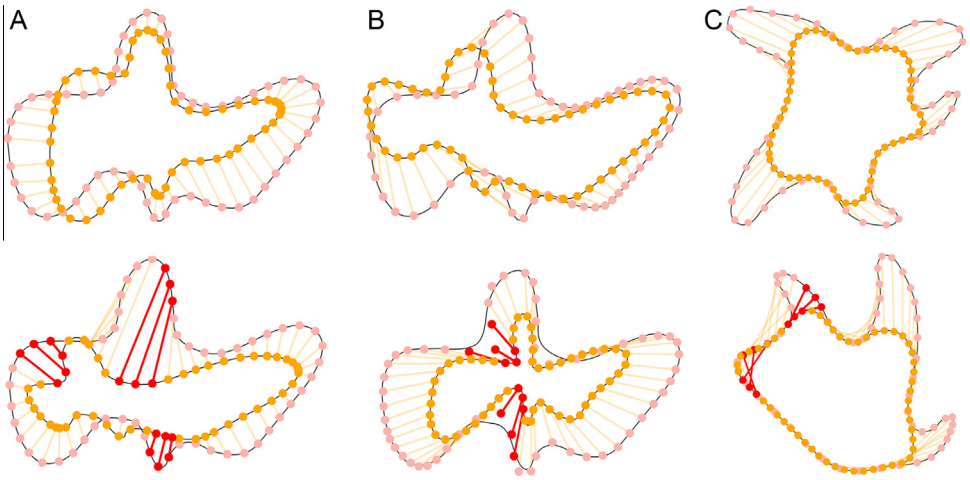


**Fig. 2.** Examples for point correspondences across non-rigid transformations established by three state-of-the-art algorithms: (A) Minimal Contour Mapping (Movahedi & Elder, 2010); (B) Coherent Point Drift (CPD) (Myronenko & Song, 2010); and (C) point set registration by preserving global and local structures (PR-GLS) (Ma et al., 2016). Transparent red dots indicate locations on the shape before transformation; orange dots, correspondences identified after transformation. Some perceptually incorrect mappings are marked in red. For (B) and (C), examples in the first row were obtained by manually fitting parameters to achieve good performance for each shape; in the second row the same parameters were used to fit a different shape, yielding errors. Examples were implemented by using publicly available code. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

achievement (El-Gaaly, Froyen, Elgammal, Feldman, & Singh, 2015; Feldman & Singh, 2006; Froyen, Feldman, & Singh, 2015). This suggests that the visual system has evolved supremely flexible mechanisms to understand even complex transformations. Here, we sought to measure in detail how human observers identify correspondence across non-rigid transformations—and define a simple model of how the visual system might achieve this.

One reason little is known about the mechanisms of perception and understanding of complex shape transformations, is that morphogenic processes are a particularly rich class of transformations. For example, all organisms change their shape with the passage of time, yielding a virtually infinite number of growth transformations. Observing growth and its driving forces is hardly possible because (i) growth is driven by unobservable mechanisms and factors within organisms, (ii) it stretches over long periods of time, and (iii) often proceeds discontinuously. Finally, growth affects different parts of an organism differently, and often alters principal features of shape, so that the appearance of an organism can change drastically across the transformation process (e.g., the metamorphosis of a tadpole into a frog). All of these factors suggest that establishing object constancy across morphogenic processes such as growth is more difficult compared to many other classes of transformations. Empirical studies on the visual perception of growth are scarce (Pittenger & Todd, 1983; Rosengren, Gelman, Kalish, & McCormick, 1991) and nonexistent for the effect of growth on visual shape representation. How might the visual system represent shape under these challenging conditions?

In his much-admired book *On Growth and Form*, Thompson (1942) suggested that the shapes of diverse and phylogenetically distant species are often related to one another by simple non-rigid transformations (Fig. 3a). This has inspired some perception researchers to suggest that the visual system may use a similar approach to represent complex shapes and shape transformations (Elder, Oleskiw, Yakubovich, & Peyré, 2013; Graf, 2006; Mark et al., 1988; Shaw & Pittenger, 1977; Todd, 1982; Todd, Weismantel, & Kallie, 2014). We put this idea to the test. We reasoned that if observers can infer non-rigid transformations—such as shearing, bloating or biological growth—they should be able to estimate how points on or near an object shift in space as a result of the transformation process. More specifically, we suggest that the ability to identify corresponding points, and track how those points shift, is one crucial perceptual component in the inference of transformations. In turn,

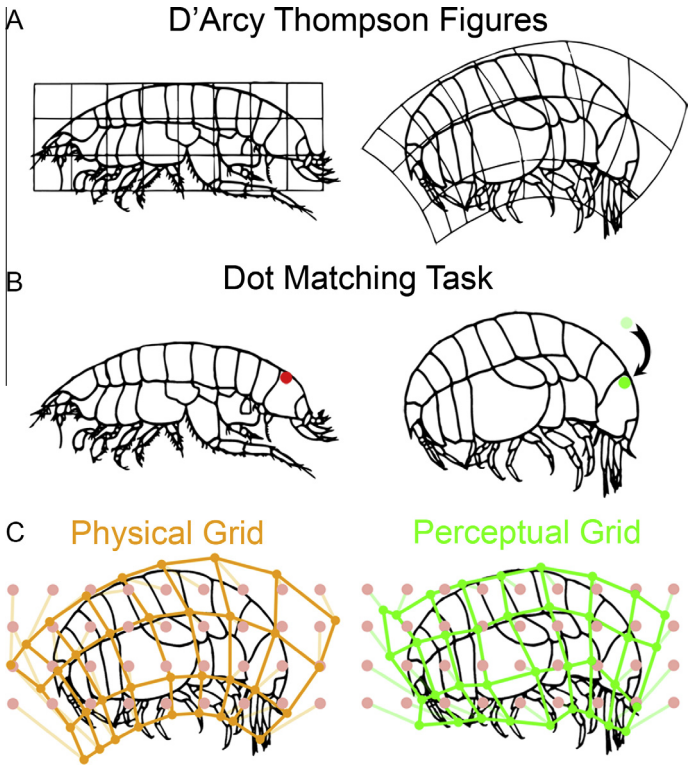


**Fig. 3.** (A) Example objects with the original transformation grid by Thompson (1942) superimposed [copyright by Cambridge University Press, reproduced with permission]. (B) A single trial of the dot matching task in which participants were asked to move a green dot to a location on the right side that they perceive as corresponding to the location of the red dot on the left side. (C) True mathematical transformation ('Physical Grid') and mean perceived correspondences ('Perceptual Grid'). Transparent red dots indicate probe locations. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

this inference could not only provide the means for estimating transformations, but also for segmenting shapes into distinct causal components: those that belong to the intrinsic identify of the shape, and those that are due to extrinsic processes applied to the object. We sought to measure—and model—how adept observers are at solving this challenging perceptual organization computation.

To test this, we showed participants pairs of objects ('before' and 'after' a transformation) and asked them to indicate corresponding points on the two objects (Fig. 3b). On each trial, a probe location was indicated by a red dot on or near the 'before' object, and the participants were asked to place a second dot "at the corresponding location" on or near the 'after' object. Note that given only these instructions the task is mathematically ill posed: there is an infinite number of possible transformations, so the points could shift to potentially arbitrary positions. The task only makes sense, and yields meaningful results, if observers identify how features from one shape transform to match those of the other. It was this process that we sought to measure. The probe points were sampled in random order from a rectangular grid (never shown to participants), or from points along the contour. Testing many of these locations enabled a dense mapping of perceived correspondences (Koenderink, Kappers, Pollick, & Kawato, 1997; Koenderink, Van Doorn, Kappers, & Todd, 1997; Moran & Leiser, 2002; Phillips, Todd, Koenderink, & Kappers, 1997, 2003).

Fig. 3c shows example settings from one of Thompson's own drawings (we removed his grids before showing them to participants; complete results are provided in Appendix A, Fig. A.1). The similarity between Thompson's grids and the responses of participants is quite striking, suggesting that

untrained observers have robust and reliable visual intuitions about which features match one another across species. The findings suggest that points both within and outside the object are perceived to shift systematically as a result of the transformation process, supporting the idea that space is represented relative to key landmarks on objects, and that perceptual organization of space can play a central role in the inference of transformations.

In these examples with recognizable objects, it is possible that participants depended heavily on familiar features, such as eyes or fins, to orient themselves. Thus, to test the generality of these results, we conducted a first systematic experiment with three unfamiliar shapes each subjected to five different non-rigid transformations.

## 2. Experiment 1

### 2.1. Material and methods

#### 2.1.1. Participants

40 students from the Justus-Liebig-University Giessen, Germany (4 male, ages 20–31), with normal or corrected vision participated in the experiment for financial compensation. All participants gave informed consent, were debriefed after the experiment, and were treated according to the ethical guidelines of the American Psychological Association. All testing procedures were approved by the ethics board at Justus-Liebig-University Giessen and were carried out in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki).

#### 2.1.2. Stimuli

Three base shapes were handcrafted with Adobe Illustrator (size 12.24°:18.35° of visual angle) and transformed by five different vector field transformations. These transformations were implemented using Matlab's built-in functions *imtransform* and *interp2* and resulted in (1) left-side compression plus a shear distortion; (2) left-side compression plus a sine bend distortion; (3) left-side barrel distortion; (4) central pincushion distortion; and a (5) central swirl distortion. All transformed images were scaled and rotated (when necessary) to obtain the same height and orientation as the base shapes; scaling and rotation were the same for all three shapes for a particular transformation. See Fig. A.2 for high resolution versions of the transformation grids and transformation vector fields. All stimuli are available for download at http://dx.doi.org/10.5281/zenodo.61701.

#### 2.1.3. Procedure

Stimuli were presented in black on a white background on a Dell U2412M monitor at a resolution of 1920 × 1200 pixels, controlled by Matlab using the Psychophysics Toolbox extension (Kleiner, Brainard, & Pelli, 2007).

On each trial, participants were presented simultaneously with one of the five base shapes (left) and one of its transformed versions (right). A probe location was indicated by a red dot (0.21°) on or near the 'before' object, and the participants were asked to place a green dot (0.21°) "at the corresponding location" on the right side (i.e., on or near the 'after' object). After each response, the probe point was replaced by the next point. Only after participants responded to all points, was another pair of shapes presented in the subsequent trial. The instructions did not refer to the shapes, and responses were allowed anywhere on the screen.

For each shape and transformation, 30 probe points were placed at equidistant positions around the perimeter of the shape, and 35 probe points were ordered in a regular 5 × 7 grid (10% bigger than the bounding box of the shape). The grid was the same for all shapes, and the probe points were the same for all participants. Neighboring points on a contour were never presented consecutively. Baseline performance was measured via trials with identical, untransformed shapes on the left and right. Each participant responded to 6 out of the 18 possible combinations of shape (3) and transformation (5 warps, 1 base shape): twice to each base shape (each time with a different transformation), and once to each transformation. Consecutive trials never featured the same shape. After completion, par-

ticipants were asked which response strategies they adopted (open format, followed by a multiple choice from a selection of strategies).

#### 2.1.4. Analysis

Four participants who explicitly reported ignoring the shape were not included in the spatial response analyses. For the grid, we calculate as a measure of accuracy the error, that is, the Euclidean distance between each response and the true location of the transformed probe point (given by the mathematical transformation that was used to generate the 'after' stimulus from the 'before' stimulus). For the contour, we first projected each response point perpendicularly onto the contour (this was typically a very short distance, as participants sought to place the dots exactly on the contour), and then computed the distance to the transformed probe point along the contour. We report this average error in degrees of visual angle and as a percentage of the perimeters of the 'after' objects, thereby giving accuracy in retinal image space and in object-centered response space (as our results show that participants respond relative to the 'after' object).

Following the same logic, accuracy of random responses (i.e., chance) was defined by the distance of responses to 10,000 bootstrapped locations within the bounding box of the grid and on the contour. Accuracy of the best rigid fit was defined by the distance of responses to the result of a Procrustes analysis allowing for rotation, reflection, scaling, and translation (projected on the contour of the 'after' object for contour probe locations). For accuracy of model predictions see below.

As a measure of participants' consistency, we defined a congruity measure by obtaining the average unsigned distance between all responses (across participants) for each probe location, and then calculating the grand average across all probe locations for each transformation.

We also computed the tendency to shift points in the direction of the transformation: this ground truth gain was obtained by projecting the perceptual vector (probe location → response location) onto the physical vector (probe location → transformed probe location) and then dividing by the length of the physical vector.

$$ground\ truth\ gain = \frac{(v_{physical} \cdot v_{perceptual})}{\|v_{physical}\|}$$

Equivalently, we computed the general willingness to shift points in space, independent of the direction of the transformation: this error gain was obtained by projecting the perceptual vector onto the orthogonal physical vector, that combines all shifts away from the transformation direction.

$$error\ gain = \frac{(v_{orth\_physical} \cdot v_{perceptual})}{\|v_{physical}\|}$$

Statistical tests were computed across shapes. We report effect size $d$, where 0.02 represents a small, 0.5 a medium, and 0.8 a large effect (Cohen, 1988).

### 2.2. Results and discussion

The 36 (out of 40) participants who reported taking the shapes into account when trying to identify corresponding locations showed strong effects of the non-linear transformations (Fig. 4; complete results are provided in Appendix A, Fig. A.3). To compare their performance to chance, we plotted their responses on a continuum between ground truth response accuracy [1] and the accuracy of random responses [0]—thus, 0.5 would mean that the average distance between responses and ground truth is the same as that between responses and random responses. As for the recognizable objects in Thompson's drawings, performance was far above chance (156.35 vs. 2362.57, [$T(70) = -405.15$, $p < .001$, $d = 94.47$]; Fig. 4d). Average error was 0.67° of visual angle (contour: 0.38°, grid: 0.93°) or 1.32% of the perimeter of the 'after' object (contour: 0.74%, grid: 1.82%). Additionally, participants tended to be highly consistent with one another; errors varied between 0.53° and 0.93° ($SD = 0.09$°). Finally, for all transformations, both errors [mean $r = -.52$; with all $r(173) < -.39$, $p < .001$] and congruity [mean $r = -.69$; with all $r(173) < -.62$, $p < .001$] dropped with the distance from the contour.

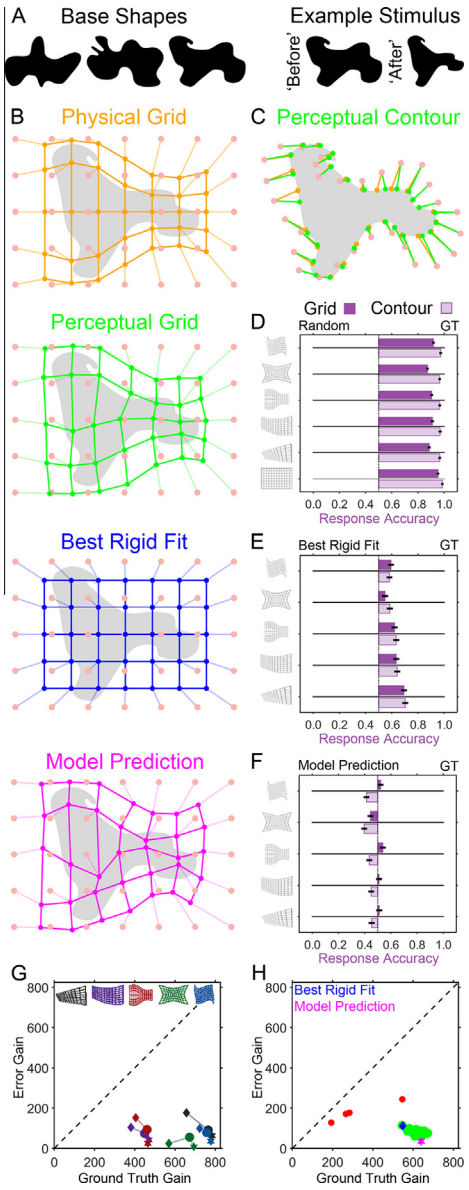


**Fig. 4.** Results of Experiment 1. (A) All 3 shapes before transformation (left) and an example shape before and after a transformation (right). (B) Mathematical transformation ('Physical Grid'); mean perceived grid correspondences ('Perceptual Grid'); best-fitting rigid grid approximation; and grid model prediction. See text for further details. Transparent red dots indicate probe locations. (C) Comparison between the mathematical transformation (orange dots) and mean perceived correspondences on the contour (green dots, 'Perceptual Contour'). (D) Mean response accuracy on a continuum between bootstrapped random responses ('Random') and the mathematical transformation (Ground Truth, 'GT') for grid (full bars) and contour probe points (transparent bars). Accuracy is calculated across shapes, and is reported separately for the baseline and each transformation. (E) Mean response accuracy on a continuum between the best-fitting rigid approximation ('Best Rigid Fit') and GT. (F) Mean response accuracy on a continuum between the predictions of our model ('Model Prediction') and GT. (G) Mean vector components in the direction of the transformation ('Ground Truth Gain') versus in the orthogonal direction ('Error gain') for individual responses (circles), the best-fitting rigid approximation (diamonds), and the model predictions (stars). (H) Ground Truth Gain and Error Gain for individual participants (circles); participants ignoring the shapes are marked in red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Because previous work showed very good performance for inferring rigid transformations (Schmidt, Spröte, & Fleming, 2016), we tested whether responses are simply a result of approximating each process with a single rigid transformation. It could be, for example, that participants are capable of identifying the best rigid approximation to the transformation (e.g., the overall orientation or scale of the object), but could not infer spatial variations that are the hallmark of non-rigid transformations. The best-fitting rigid approximation to one example transformation is shown in Fig. 4b. Participants' responses were on average significantly closer to the true non-rigid locations than those predicted by Procrustes analysis (156.35 vs. 238.83, $[T(70) = -18.39, p < .001, d = 4.29]$; Fig. 4e), suggesting that participants did indeed take non-rigid aspects into account.

When plotting ground truth against error gain, we can evaluate to what extent responses were shifted in the direction of the transformation (area below diagonal) versus in other directions (area above diagonal). Responses are clearly shifted into the direction of the transformation, and more so than the best-fitting rigid approximation (Fig. 4g). Finally, even the four participants that reported trying to ignore the shape (red dots), were nevertheless affected by the transformations (Fig. 4h).

### 2.2.1. Modeling

As noted in the introduction, identifying correspondence across arbitrary non-rigid transformations is computationally challenging: any given point in space near the 'before' object could in principle map to any other point in space surrounding the 'after' object. Thus, to compute correspondences, it seems likely that the brain makes certain simplifying assumptions, such as that the transformations are spatially smooth, rather than discontinuous. We suggest that the visual system identifies and establishes correspondence for a few salient landmarks on the shape, and infers the position of other points in space relative to these. This would provide a robust method to infer correspondence.

We define salient landmarks by local maxima of contour 'surprisal'—an information theoretic measure, related to curvature, which indicates how much each point on the contour 'stands out' from its local context (Attneave, 1954; see Feldman & Singh, 2005, for derivation). Specifically, surprisal makes use of the assumption that contours are most likely to continue along their current tangent direction (Elder & Goldberg, 2002; Field, Hayes, & Hess, 1993; Geisler, Perry, Super, & Gallogly, 2001; Grossberg & Mingolla, 1985). Consequently, the most informative (i.e., least predictable) points on the contour are those that diverge from this current direction, therefore 'standing out' from their local context. This can be formalized by a continuous probability distribution (von Mises distribution) on the turning angles, centered on 0—resulting in monotonically decreasing probabilities ($p$) with increasing divergence from the current tangent direction. Surprisal is formalized as $u = -\log(p)$, and therefore increases with the magnitude of turning angle (Feldman & Singh, 2005).

First, we measured how performance varied as a function of surprisal. Based on previous findings, we calculated turning angles using a window integration size of 5% of the contour perimeter (Schmidt et al., 2016). We treated positive and negative curvature (i.e., convex and concave contour segments) symmetrically (*unsigned*; Attneave, 1954; Norman, Phillips, & Ross, 2001) and normalized the surprisal for each point on a contour with respect to the maximum surprisal on that contour. Consistent with previous findings with rigid transformations (Phillips et al., 1997; Schmidt et al., 2016), we find that for all transformations the accuracy of participant's settings for the points on the contour correlated, albeit somewhat weakly, with contour surprisal [mean $r = .12$; significant in 31 participants with $r(148) > .16, p < .042$]. This provides an initial indication, as should be expected, that easily localizable features (e.g., corners) yield more precise estimates of spatial transformations. However, the weakness of the correlation also indicates that performance is not limited by the ability of participants to localize points on the contour based solely on their local characteristics. In other words, participants do not localize and distinguish points on the boundary based simply on their local curvatures: points on quite straight (i.e., low surprisal) contour regions can also be localized precisely, suggesting participants use non-local features to determine position.

Second, as a more thorough test of the hypothesis, we implemented a simple model for predicting the perceived locations of transformed points by interpolating and extrapolating their positions relative to a small number of salient landmarks on the contours of the 'before' and 'after' shapes.

Landmarks were identified by (i) calculating the normalized distribution of surprisal values along the contour of the base shape (−1, 1) as described above and (ii) finding local minima/maxima that are surrounded by values that are higher/lower by 0.05 on both sides and have absolute values >0.02. Then, we projected all tested probe points onto the contour of the 'before' shape and defined their positions relative to their neighboring landmarks. Specifically, by finding the same relative position between the corresponding two landmarks on the 'after' shape, we can estimate corresponding points after transformation. For example, if a probe point lay one third of the distance between two landmarks on the 'before' shape, its match was predicted to lie one third of the distance between the two corresponding landmarks on the 'after' shape. Points from the grid were then projected back along the contour normal by a distance proportional to the difference between the shape perimeters (for illustration see Fig. 5). This produced a predicted response location for every probe point that we tested.

We find that on average, the predicted locations were even closer to the participants' settings than the ground truth locations (122.48 vs. 156.35, [$T(70) = 7.88$, $p < .001$, $d = 1.84$]; Fig. 4f) and that the pattern of the residual errors in the model are similar to the pattern of participants' responses (Fig. 4g). Average distance (i.e., error) between the model and participant's responses was 0.53° of visual angle (contour: 0.24°, grid: 0.78°) or 1.12% of the perimeter of the 'after' object (contour: 0.50%, grid: 1.66%). This suggests that the model accurately predicts both the good overall performance and the specific pattern of errors that participants made. Importantly, this is a zero-free-parameter model: no fitting of the model to the data was performed. Finally, we found that the model performs worse when based

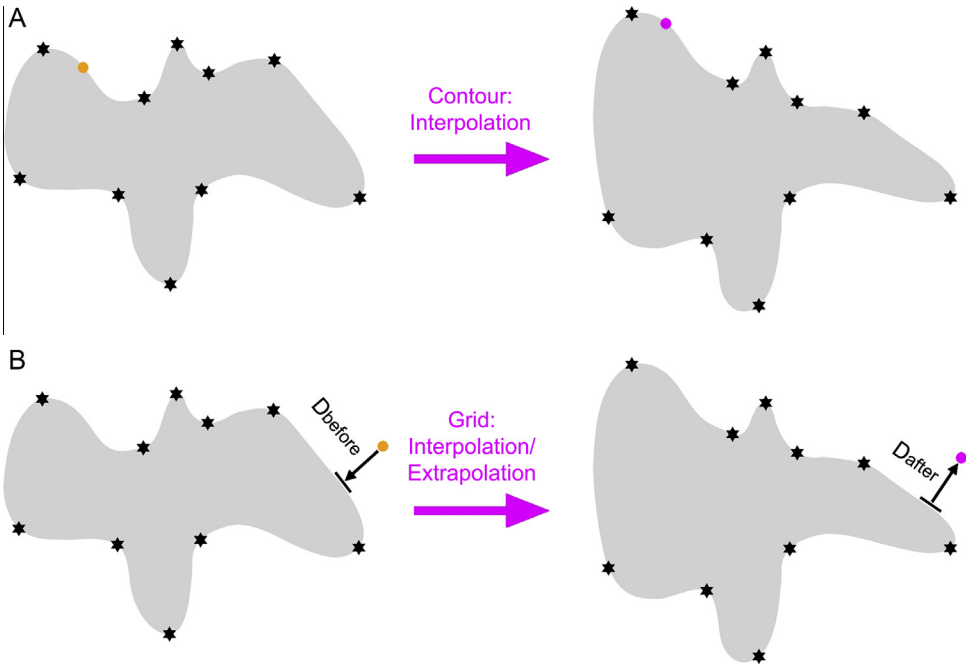


**Fig. 5.** Illustration of our model. (A) Prediction of contour responses based on interpolation between salient parts of the contour. First, we project each probe point onto the contour of the 'before' object and define its position relative to neighboring landmarks (stars). Second, we predict its position after transformation by interpolating between the corresponding landmarks on the 'after' object. (B) Prediction of grid responses based on interpolation and extrapolation. First, we project each probe point onto the contour of the 'before' object and use the same interpolation process as described in (A). Second, we project the new location on the 'after' object back along the contour normal. For global transformations that affect the whole shape (Experiment 1), this projection distance ($D_{after}$) is weighted by the difference between the perimeter ($P$) of the 'before' and 'after' object; $D_{after} = D_{before} * (P_{after}/P_{before})$. For local transformations (Experiment 2), the projection distance is not weighted; $D_{after} = D_{before}$.

on points of inflection rather than on local maxima of surprisal (150.29 vs. 122.48, [$T(70) = -8.85$, $p < .001$, $d = 2.06$]), supporting our suggestion (based on Attneave, 1954 and Feldman & Singh, 2005), that salient landmarks play a privileged role in the computations.

## 3. Experiment 2

The transformations in Experiment 1 are smooth, global deformations of shape. However, natural non-rigid transformations, such as growth processes, are not always global; many processes are restricted to specific parts of objects, such as limbs. In Experiment 2, we sought to test whether the ability to infer non-rigid transformations generalized to such spatially heterogeneous processes (Denisova, Feldman, Su, & Singh, 2016) by using shapes in which limb-like protuberances emerged out of the objects as a result of the transformation.

### 3.1. Material and methods

#### 3.1.1. Participants

18 students from the Justus-Liebig-University Giessen, Germany (5 male, ages 19–34) participated in the experiment, for further details see Experiment 1.

#### 3.1.2. Stimuli

Five base shapes were handcrafted with Adobe Photoshop (size of bounding box: 14.50°:13.20°, 19.04°:11.40°, 12.26°:15.37°, 12.34°:18.76°, and 12.12°:13.63° of visual angle) and transformed by using the Liquify tool of Adobe Photoshop® to stretch out convexities in the base shapes into limb-like protuberances. All stimuli and transformations meshes are available for download at http://dx.doi.org/10.5281/zenodo.61701.

#### 3.1.3. Procedure

The procedure was that same as in Experiment 2, except that 40 probe points were placed at equidistant positions around the perimeter of the shape, and the grid was a regular $5 \times 5$ grid of 25 probe points. Each participant responded to all 5 shapes and to 1 of the base shapes, and was asked for adopted response strategies after completion.

#### 3.1.4. Analysis

Three participants who explicitly reported ignoring the shape were not included in the spatial response analyses. Accuracy was computed as in Experiment 1; ground truth was based on the mesh transformation of the Liquify tool. Since most of the probe points were not affected by the transformations, ground truth gain, error gain, and Procrustes analysis were not informative and were not calculated. Note that model predictions for points from the grid were obtained by projecting back along the contour normal by the same distance as they were projected onto the contour; as transformations were local the model did not need a global weighting factor. Statistical tests were computed across shapes.

### 3.2. Results and discussion

As with global transformations, we find that the 15 (out of 18) participants who reported taking the shapes into account in their judgments were extremely good at establishing correspondence (Fig. 6; complete results are provided in Appendix A, Fig. A.4)—and far closer to ground truth than chance would predict (155.95 vs. 1888.25, [$T(28) = -206.72$, $p < .001$, $d = 73.44$]; Fig. 6d). Average error was 1.34° of visual angle (contour: 1.01°, grid: 1.88°) or 1.81% of the perimeter of the 'after' object (contour: 1.35%, grid: 2.54%); individual participants varied between 0.93° and 1.90° ($SD = 0.28°$). Again, errors [mean $r = -.26$; significant in 12 participants with $r(123) < -.18$, $p < .004$] and congruity [$r(123) = -.59$, $p < .004$] dropped with distance to the contour. The model based on interpolation between salient landmarks, defined by surprisal equivalent to Experiment 1, predicted the responses

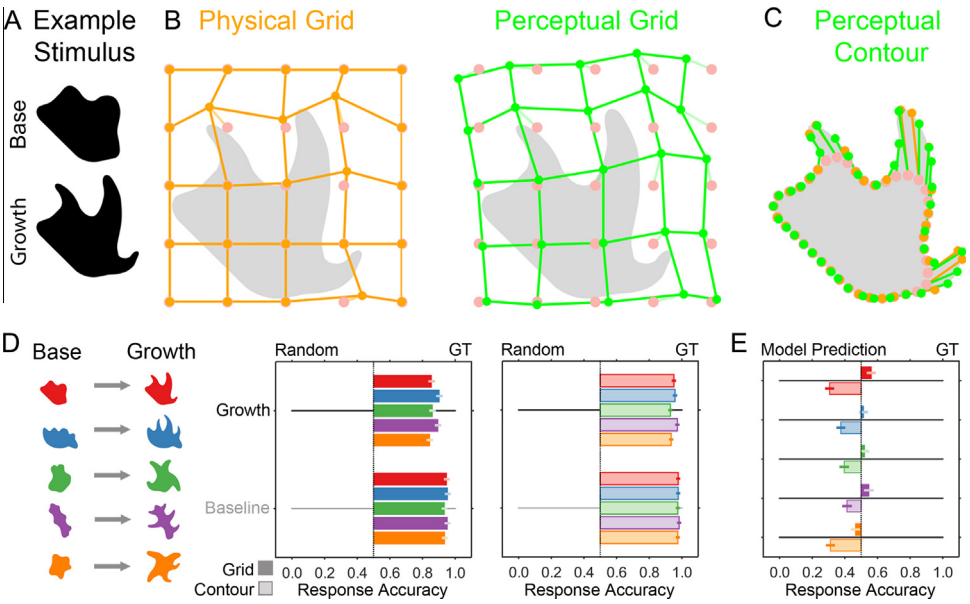


**Fig. 6.** Results of Experiment 2. (A) An example shape before and after the growth transformation. (B) Mathematical transformation ('Physical Grid') and mean perceived grid correspondences ('Perceptual Grid'). Transparent red dots indicate probe locations. (C) Comparison between the mathematical transformation (orange dots) and mean perceived correspondences on the contour (green dots, 'Perceptual Contour'). (D) Mean response accuracy on a continuum between bootstrapped random responses ('Random') and the mathematical transformation (Ground Truth, 'GT') for grid (full bars) and contour probe points (transparent bars). Accuracy is calculated per shape (different colors) and reported separately for the baseline and the growth transformation. (E) Mean response accuracy on a continuum between the predictions of our model ('Model Prediction') and GT for grid (full bars) and contour probe points (transparent bars). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

better than ground truth for contour (44.92 vs. 116.77, [$T(28) = 13.24$, $p < .001$, $d = 4.71$]) but not for grid locations (278.18 vs. 218.64, [$T(28) = -14.59$, $p < .001$, $d = 5.18$]; Fig. 6e). Average distance (i.e., error) between the model and participant's responses was 1.16° of visual angle (contour: 0.39°, grid: 2.40°) or 1.56% of the perimeter of the 'after' object (contour: 0.52%, grid: 3.23%).

This suggests, again, that for contour locations participants orient themselves using salient features on the boundary, and infer the positions of transformed points relative to these. Note however that neither ground truth nor the model do particularly well for grid locations, suggesting that for transformations of object parts the visual system relies on more sophisticated mechanisms than simple geometrical extrapolation (e.g., segmentation and weighting of moving and non-moving parts of the contour).

## 4. Experiment 3

The heterogeneous transformations of Experiment 2 were intended to approximately resemble certain natural growth processes, but they were created in image editing software rather than through accurate physical or biological simulation. This has the advantage that we have precise control over—and knowledge of—the ground truth correspondences. However, it is also interesting to ask the extent to which observers can track more complex, realistic growth processes that can radically alter structure, such as the unfurling of a growing stem, or the development from froglet to frog. We addressed this question in Experiment 3 by using shapes derived from images of real-world organisms at different stages of development.

### 4.1. Material and methods

#### 4.1.1. Participants

15 students from the Justus-Liebig-University Giessen, Germany (6 male, ages 21–35) participated in the experiment, for further details see Experiment 1.

#### 4.1.2. Stimuli

Five base and transformed shapes were derived from photographs or drawings of real-world organisms at different stages of development: a turtle embryo, a frog, a dog, a leaf, and a fern (size of bounding box: 10.24°:16.22°, 13.21°:20.28°, 11.60°:9.90°, 12.16°:10.99°, and 17.70°:18.02° of visual angle). All stimuli are available for download at http://dx.doi.org/10.5281/zenodo.61701.

#### 4.1.3. Procedure

Same procedure as in Experiment 2, just that the grid was 110% of each base shape's bounding boxes. Each participant responded to all 5 shapes, and was asked for adopted response strategies after completion.

#### 4.1.4. Analysis

With no ground truth available, the average distance between responses (congruity) was used as response variable. Note that for random responses the congruity was defined as the average distance of random responses to the centroid of participants' responses. Statistical tests were computed across shapes.

### 4.2. Results and discussion

Although we have no means to establish the ground truth transformations, we can nevertheless measure how consistent participants are in their solution of the correspondence problem (Fig. 7; complete results are provided in Appendix A, Fig. A.5).

It is important to note, that in some cases, there is no well-defined solution for all probes. For example, where a new limb is created or destroyed, correspondence cannot be defined. Despite this, we find that overall participants were extremely consistent in their responses, certainly far above chance (average distance between responses: 88.22 vs. 1217.42, $[T(648) = -57.91$, $p < .001$, $d = 4.54]$; Fig. 7d). We also find that congruity declined as a function of the distance from the contour $[r(123) = -.66, p < .001]$. For the model, we defined salient landmarks in the 'before' object equivalent to Experiment 1 and—because we lacked ground truth—chose the corresponding landmarks in the 'after' object by hand. Model prediction was comparable to that obtained in Experiment 2 and better for contour compared to grid responses (Fig. A.6). Average distance (i.e., error) between the model and participant's responses was 0.89° of visual angle (contour: 0.42°, grid: 1.63°) or 1.66% of the perimeter of the 'after' object (contour: 0.78%, grid: 3.06%).

## 5. General discussion

Together, our findings suggest that observers are extremely good at tracking how positions in space—both on and nearby objects—are affected by complex non-rigid transformations as occur in morphogenesis. This ability to identify corresponding features, and track how those features shift as a result of a transformation, is one crucial perceptual component in the visual inference of transformations. It provides a general mechanism for representing potentially arbitrarily complex transformations, like those during viscous flow, crystal growth or biological development.

Perceiving and interacting with objects in our visual environment depends crucially on our ability to establish object constancy across transformations and to understand the nature of these transformations. For example, in biological growth, this ability is the basis for our judgments on age, kinship, or the past and future development of organisms, including other humans. Although all biological organisms are subject to growth and change their shape with the passage of time, the

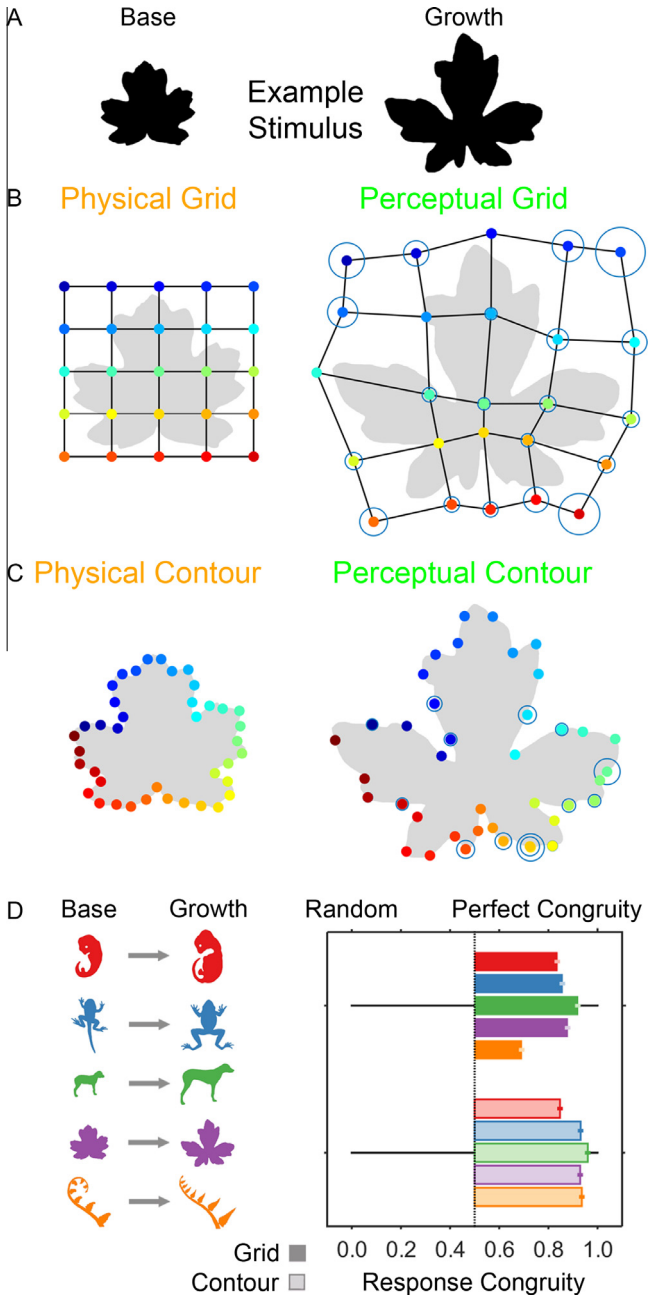


**Fig. 7.** Results of Experiment 3. (A) An example shape before and after the growth transformation. (B) Comparison between grid probe points ('Physical Grid') and the mean perceived grid correspondences ('Perceptual Grid'). The diameter of the blue circles shows the average distance between responses. (C) Comparison between contour probe points ('Physical Contour') and the mean perceived contour correspondences ('Perceptual Contour'). (D) Mean congruity on a continuum between bootstrapped random responses ('Random') and zero variance ('Perfect Congruity') for grid (full bars) and contour probe points (transparent bars). Accuracy is calculated per shape (different colors). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

visual perception of these changes has not received much attention (although see Cutting, 1978; Mark et al., 1988; Pittenger & Shaw, 1975; Pittenger & Todd, 1983).

Many previous studies have demonstrated our ability to estimate and compensate for mathematically simple transformations, such as rotations in 2D (Moran & Leiser, 2002; Schmidt et al., 2016; Shepard & Cooper, 1982) and 3D (Cornelis, Van Doorn, & Wagemans, 2009; Koenderink, Kappers et al., 1997; Koenderink, Van Doorn et al., 1997; Phillips et al., 1997; Todd, 1984). However, our findings confirm that this is a special case of a much more general ability to infer and interpret processes that alter the shape itself (Arnheim, 1974; Cutting, 1982; Cutting & Kozlowski, 1977; Graf, 2006; Kim, Effken, & Shaw, 1995; Kourtzi & Shiffrar, 2001; Leyton, 1989, 2012; Pinna, 2010; Pittenger & Shaw, 1975; Shaw & Pittenger, 1977; Spröte & Fleming, 2013), and not simply the projection of the shape on the retina.

In the visual system, processes of perceptual grouping and figure-ground organization combine image features into representations of shape and determine whether they belong to object or background (Wagemans et al., 2012). Our findings suggest that the resulting shape representations are subject to further perceptual organization processes, in which salient shape features are identified and used to establish inferences about the shape-forming, morphogenic processes that gave an object its present shape. For smooth, global transformations (Experiment 1), these inferences can be approximated by our simple model—even for probe locations away from the contour. For heterogeneous, local transformations (Experiments 2) and for complex shapes and transformations that remove salient shape features (Experiment 3), the perceptual organization processes that are used to estimate effects of transformations outside the contour seem to be more complex, probably involving segmentation and weighting of transformed and non-transformed parts of the contour.

Again, the ease with which we perform the task makes it difficult to appreciate how computationally challenging it is—note that even state-of-the-art computer algorithms are not perfect (Fig. 2) and are not designed to infer positions outside the contour when provided with contour information only.

Although previous studies have suggested that shapes may be represented and compared in terms of transformations (Atit et al., 2013; Bedford & Mansson, 2010; Chen & Scholl, 2015; Cutting, 1982; Dubinskiy & Zhu, 2003; Feldman, 1995; Feldman & Singh, 2006; Feldman et al., 2013; Hoffman & Richards, 1984; Koffka, 1935/1965; Leyton, 1989, 1992; Mark et al., 1988; Mark & Todd, 1985; Ons & Wagemans, 2011; Pittenger & Shaw, 1975; Shaw & Pittenger, 1977), to our knowledge we provide the first detailed empirical evidence that the human visual system represents positions in space on and around objects relative to salient landmarks of the object itself. For example, Leyton (1989, 1992) has developed a theory for inferring causal history (i.e. a series of transformations) from single shapes. He suggested that all curvature extrema along a shape's outline have local symmetry axes leading to them that indicate the directions along which forces operated to create the observed shape ("process inferring symmetry axis"). However, empirical evidence supporting this idea has been scarce (Leyton, 1986a, 1986b), and his model does not predict transformations of points outside objects, nor the components of transformation that are orthogonal to the medial axes of the shape.

There are also other domains, such as motor control in which perceived space is affected by allocentric scene properties (Medendorp, 2011), or even action itself. For example, numerous studies have indicated that visual space is distorted peri-saccadically (Binda, Cicchini, Burr, & Morrone, 2009). However, unlike peri-saccadic remapping—which is essentially unconscious—in our experiments, participants reported an explicit awareness of the transformation. We suggest this reflects processes for parsing and understanding object shape—and its variations—rather than involuntary effects of mapping between different reference frames (e.g., before and after saccades).

The finding that observers also re-map points that are located in free space (i.e., not on or within the object itself) suggests that the visual system represents unstructured locations in space relative to landmarks on shapes. In turn, this ability potentially provides the means not only to determine correspondences, but also to abstract transformations from shapes, that is, to recognize and estimate parameters of shape-distorting transformations, as long as they preserve salient landmarks. Thus, the tendency to represent space relative to shape features, could provide the basis for segmenting

shapes into distinct causal components: those that belong to the intrinsic identify of the shape (e.g., salient landmarks), and those that are due to extrinsic, forces, processes and events applied to the object (Arnheim, 1974; Leyton, 1989; Pinna, 2010; Spröte & Fleming, 2013). We have shown previously that observers can identify and match certain non-rigid transformations, such as bends, across a range of different objects (Spröte & Fleming, 2016). The current findings hint at some of the perceptual organization processes underlying such abilities.

More generally, the ability to infer and interpret non-rigid transformations is potentially important for a wide range of perceptual and cognitive tasks, including (i) object constancy—the ability to identify objects across diverse viewing conditions or organisms across growth; (ii) predicting motor affordances and future behavior of objects based on the ways they respond to shape-transforming forces (Battaglia, Hamrick, & Tenenbaum, 2013; Bouman, Xiao, Battaglia, & Freeman, 2013; Cholewiak, Fleming, & Singh, 2013; Paulun, Kawabe, Nishida, & Fleming, 2015; Vrins, de Wit, & van Lier, 2009); (iii) the ability to identify similarities between different shapes (Hahn, Chater, & Richardson, 2003; Hahn, Close, & Graf, 2009; Imai, 1977; Kimia, Tannenbaum, & Zucker, 1995; Kubilius, Bracci, & Op de Beeck, 2016; Ons & Wagemans, 2012; Panis, Vangeneugden, & Wagemans, 2008); and (iv) the prediction of plausible variations of objects belonging to a given class when presented with only a single, or small number of exemplars (Feldman, 1992, 1997; Fleming, 2015; Richards, Feldman, & Jepson, 1992). Thus, we suggest that the hitherto under-appreciated perception of non-rigid shape transformations is a crucial element of a broader set of 'Shape Understanding' abilities, which parse and interpret the causally significant features of shapes (Leyton, 1989; Scholl & Tremoulet, 2000; Spröte & Fleming, 2013).

In sum, our findings reveal the operation of specific perceptual organization processes that make us remarkably adept at identifying correspondences across complex shape-transforming processes. We suggest this ability is invaluable for many tasks that involve 'making sense' of shape.

## Acknowledgments

This research was funded by the DFG funded Collaborative Research Center on "Cardinal Mechanisms of Perception" (SFB-TRR 135). The authors wish to thank Bart Anderson, Steven A. Cholewiak, James Elder, Andrew Welchman, and Patrick Spröte for helpful comments, and Jiayi Ma and Viva Movahedi for help in implementing their algorithms for Fig. 2. Thanks to Dejan Todorovic for suggesting to use D'Arcy Thompson's drawings as stimuli and to Eugen Prokott for data collection and Harun Karimpur for data collection and stimulus pre-processing. Individual observer response data and stimuli for all experiments, together with transformation mesh files for Experiment 2 are available at http://dx.doi.org/10.5281/zenodo.61701.

## Appendix A

See Figs. A.1–A.6.

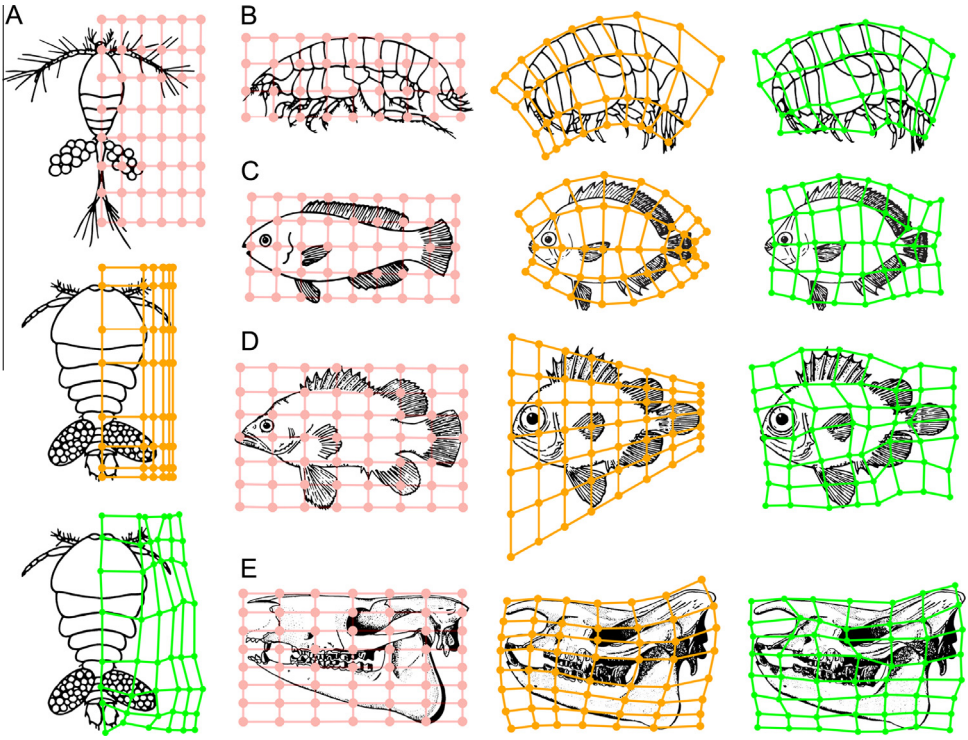


**Fig. A.1.** Test stimuli and results for the five object pairs derived from original drawings by Thompson (1942) [copyright by Cambridge University Press, reproduced with permission]: (A) copepods, (B) amphipods, (C) Parrotfish/Angle fish, (D) Wreckfish/Short Bigeye, and (E) Rhinoceros skulls. For each pair, we show the 'before' object with the grid of probe locations in transparent red, the 'after' object with the transformed grid in orange, and the 'after' object with the mean responses in green. The size of the bounding box of the 'before' object was 17.68°:20.24°, 21.14°:9.60°, 20.43°:10.00°, 20.77°:13.22°, and 20.45°:11.93°, respectively. In A, C, and D, accuracy is clearly decreasing with distance from the contour. Also, the significance of shape features is evident. For example, in D participants show adequate reproduction of ground truth only within the shape. Outside the shape, they respond with respect to the features of the fish (e.g., by interpolating between dorsal and tail fin). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

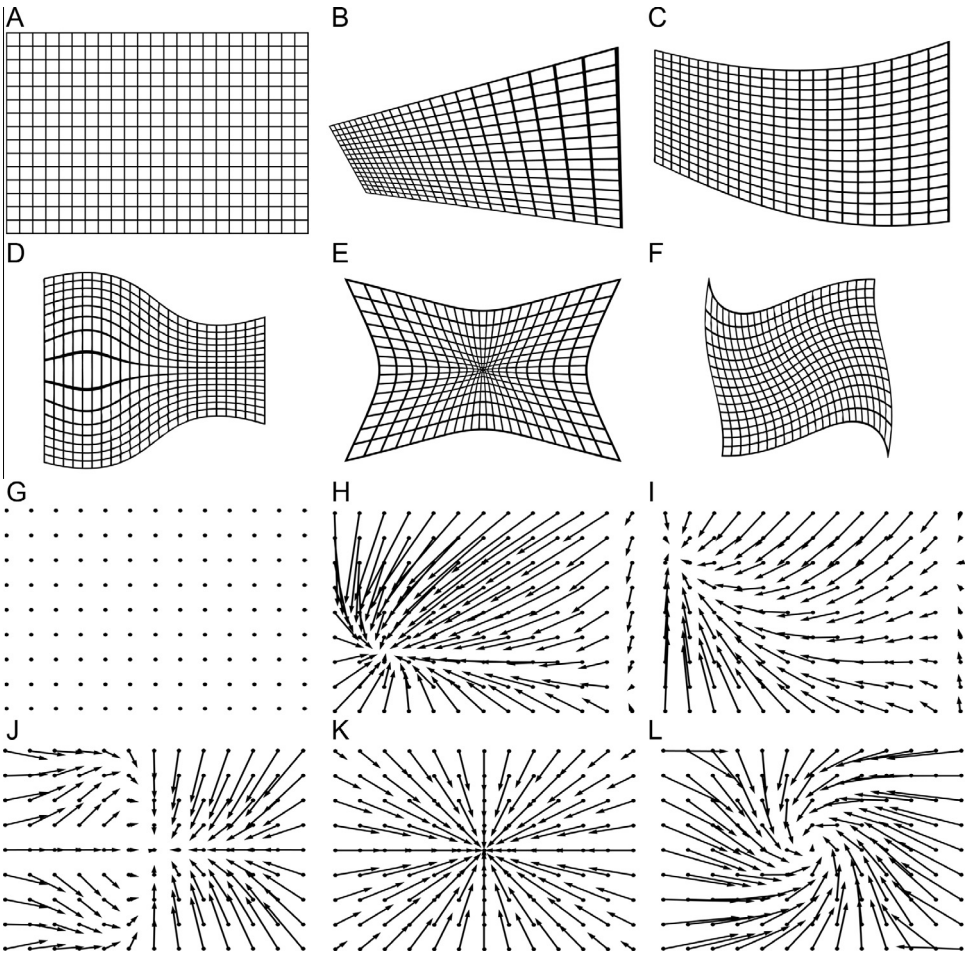


**Fig. A.2.** High resolution versions of the five transformation grids (B)–(F) and corresponding transformation vector fields (H)–(L) that were used to generate stimuli of Experiment 1. Transformation grid and vector field of the baseline is given in (A) and (G) for comparison. Note that the mathematical transformations were continuous; the number of grid rows and columns and of transformation vectors were chosen for illustration purposes only.

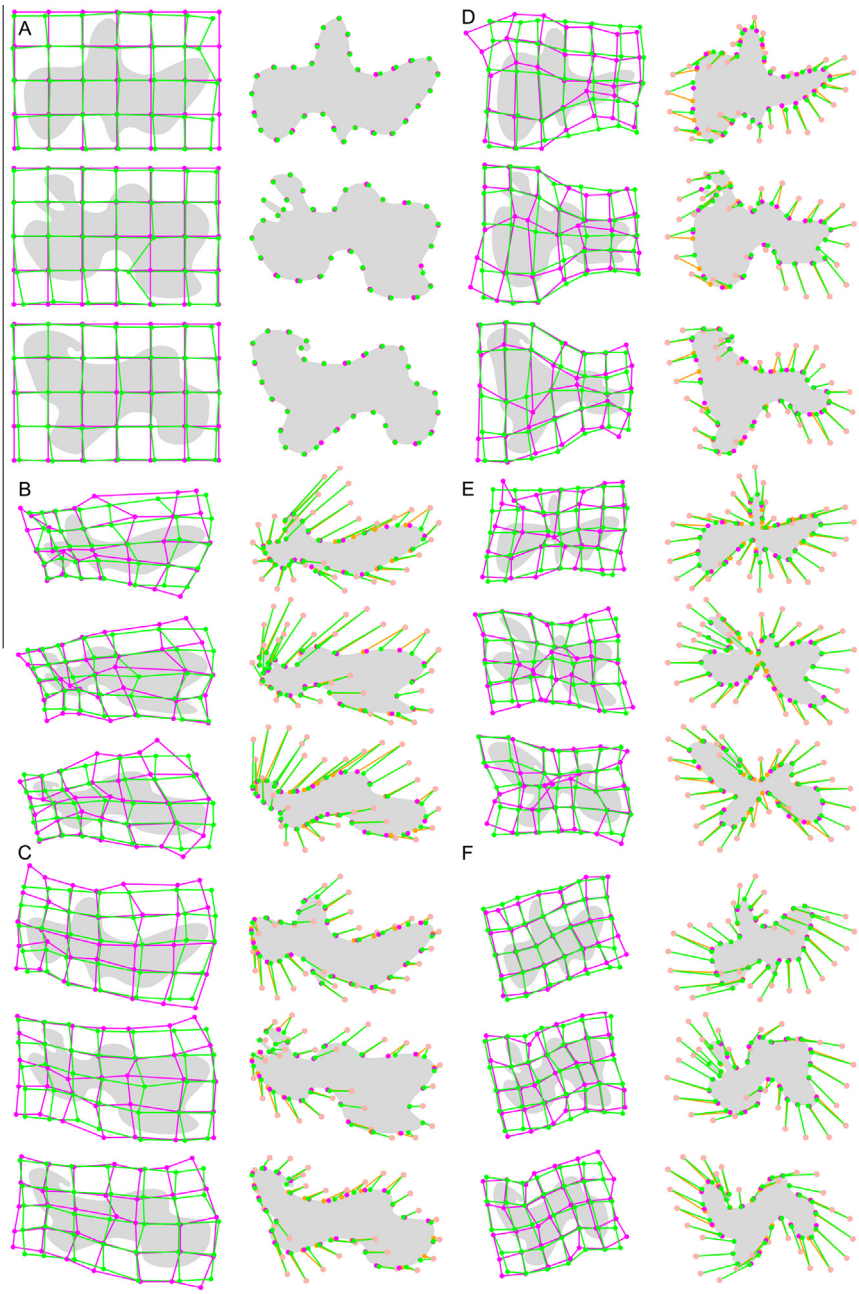


**Fig. A.3.** Results and model predictions for baseline (A) and the five transformations (B)–(F) of Experiment 1. For each transformation, we show the 'after' object of each shape; for grid probe locations (left) we show mean responses in green and model predictions in purple; for contour probe locations (right) we show probe locations in transparent red, mean responses in green, ground truth transformation in orange, and model predictions in purple. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

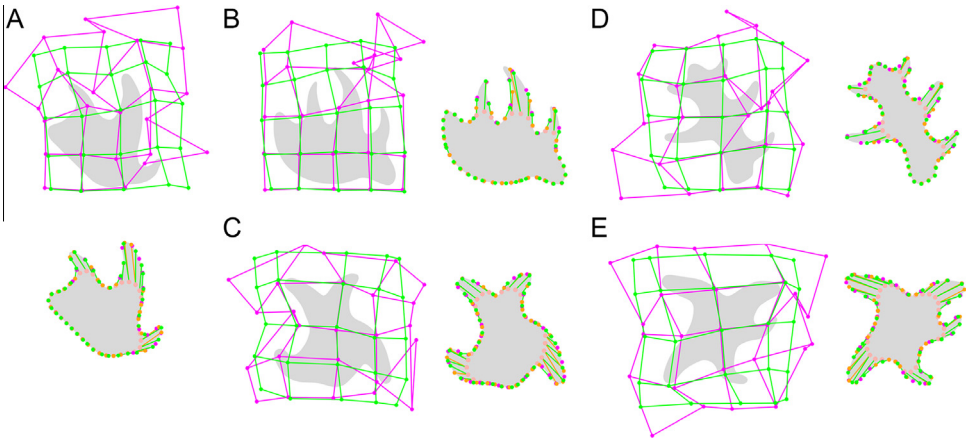


**Fig. A.4.** Results and model predictions for each shape (A)–(E) of Experiment 2. For each shape, we show the 'after' object; for grid probe locations we show mean responses in green and model predictions in purple; for contour probe locations we show probe locations in transparent red, mean responses in green, ground truth transformation in orange, and model predictions in purple. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

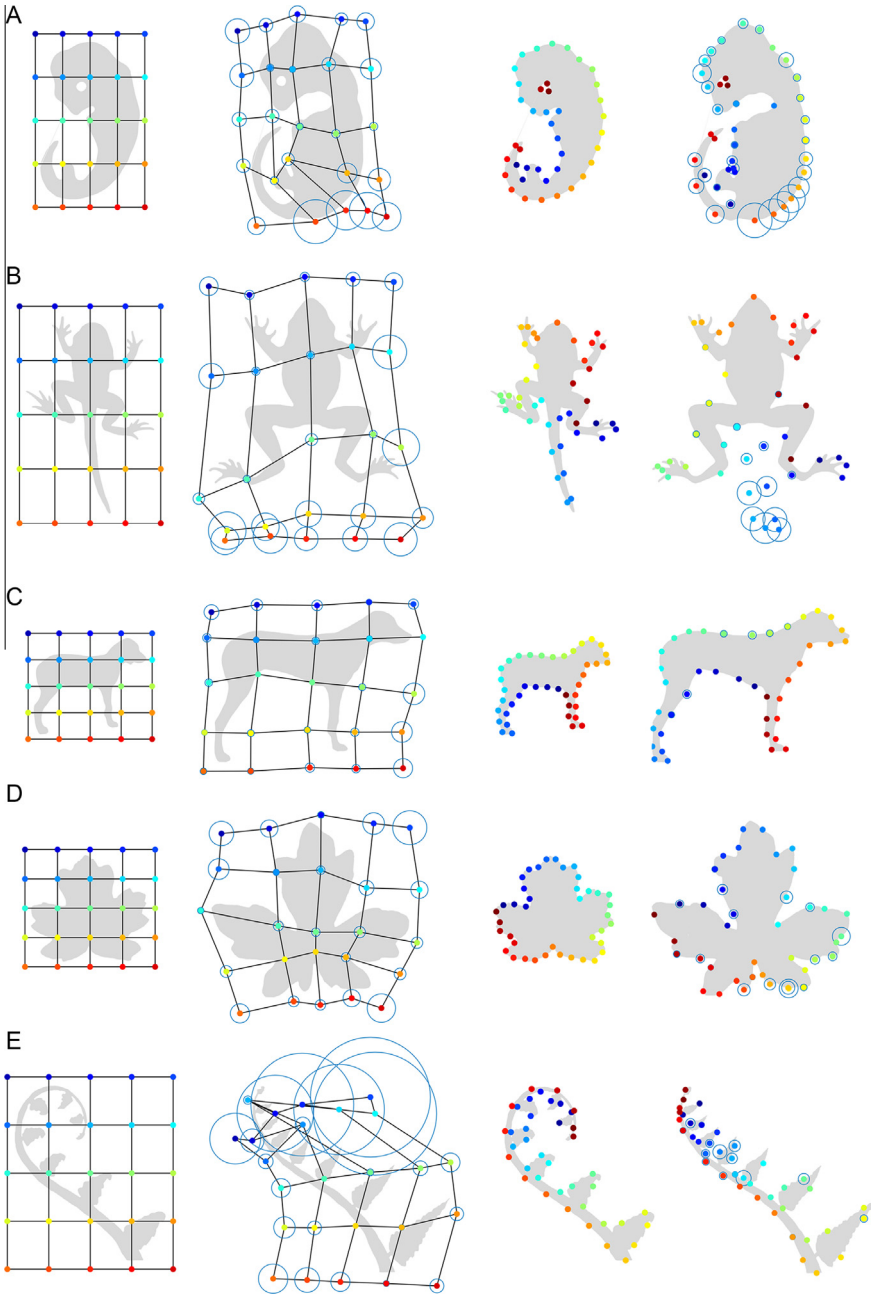


**Fig. A.5.** Probe locations and results for each object pair (A)–(E) of Experiment 3. For each pair, we show from left to right: the 'before' object with grid probe locations, the 'after' object with mean responses, the 'before' object with contour probe locations, and the 'after' object with mean responses. The diameter of the blue circles shows the average distance between responses. Object pairs were based on (A) drawings of a turtle embryo by Haeckel (1874), (B) photographs of a froglet [Eric Isselee/Shutterstock.com] and an adult common frog [yyang/Shutterstock.com], (C) a time-lapse video of a growing Rhodesian Ridgeback [copyright 2014 by Greg Coffin], (D) photographs of a baby and an adult field maple leaf [copyright 2015 by first author], and (E) drawings of a silver fern [copyright 2011 by Jane Maisey; designjane.com]. Permission was obtained from all copyright holders. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

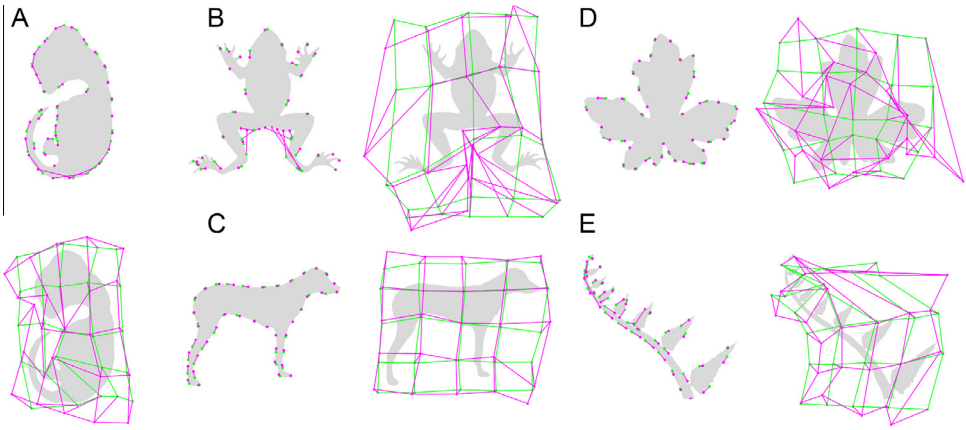


**Fig. A.6.** Results and model predictions for each shape (A)–(E) of Experiment 3. For each shape, we show the 'after' object; for grid and contour probe locations we show mean responses in green and model predictions in purple. Note, how the model fails when salient features of the contour are changed by the growth transformation ('frog tail', B) or when the contour is very detailed (field maple leaf, D). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

## References


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

Atit, K., Shipley, T. F., & Tikoff, B. (2013). Twisting space: Are rigid and non-rigid mental transformations separate spatial skills? *Cognitive Processing, 14*, 163–173.

Attneave, F. (1954). Some informational aspects of visual perception. *Psychological Review, 61*(3), 183–193.

Battaglia, P. W., Hamrick, J. B., & Tenenbaum, J. B. (2013). Simulation as an engine of physical scene understanding. *Proceedings of the National Academy of Sciences, 110*, 18327–18332.

Bedford, F. L., & Mansson, B. E. (2010). Object identity, apparent motion, transformation geometry. *Current Research in Psychology, 1*, 35–52.

Binda, P., Cicchini, G. M., Burr, D. C., & Morrone, M. C. (2009). Spatiotemporal distortions of visual perception at the time of saccades. *The Journal of Neuroscience, 29*, 13147–13157.

Boettiger, A., Ermentrout, B., & Oster, G. (2009). The neural origins of shell structure and pattern in aquatic mollusks. *Proceedings of the National Academy of Sciences, 106*, 6837–6842.

Bouman, K. L., Xiao, B., Battaglia, P., & Freeman, W. T. (2013). Estimating the material properties of fabric from video. In *Proceedings of the 19th IEEE international conference on computer vision* (pp. 1984–1991).

Carlson, B. M. (2013). *Human embryology and developmental biology*. Elsevier Health Sciences.

Chen, Y. C., & Scholl, B. (2015). The perception of history: Seeing causal history in static shapes is powerful enough to induce illusory motion perception. *Journal of Vision, 15*(12). 1035-1035.

Chen, D. T., Wen, Q., Janmey, P. A., Crocker, J. C., & Yodh, A. G. (2010). Rheology of soft materials. *Annual Review of Condensed Matter Physics, 1*, 301–322.

Cholewiak, S. A., Fleming, R. W., & Singh, M. (2013). Perception of physical stability and center of mass of 3-D objects. *Journal of Vision, 15*, 13.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.

Cornelis, E., Van Doorn, A. J., & Wagemans, J. (2009). The effects of mirror reflections and planar rotations of pictures on the shape perception of the depicted object. *Perception, 38*, 1439–1466.

Cutting, J. E. (1978). Perceiving the geometry of age in a human face. *Attention, Perception, & Psychophysics, 24*, 566–568.

Cutting, J. E. (1982). Blowing in the wind: Perceiving structure in trees and bushes. *Cognition, 12*, 25–44.

Cutting, J. E., & Kozlowski, L. T. (1977). Recognizing friends by their walk: Gait perception without familiarity cues. *Bulletin of the Psychonomic Society, 9*, 353–356.

Denisova, K., Feldman, J., Su, X., & Singh, M. (2016). Investigating shape representation using sensitivity to part- and axis-based transformations. *Vision Research, 126*, 347–361.

Dubinskiy, A., & Zhu, S. C. (2003). A multi-scale generative model for animate shapes and parts. In *Proceedings of the 9th IEEE international conference on computer vision* (pp. 249–256).

Elder, J. H., & Goldberg, R. M. (2002). Ecological statistics of Gestalt laws for the perceptual organization of contours. *Journal of Vision, 2*, 5.

Elder, J. H., Oleskiw, T. D., Yakubovich, A., & Peyré, G. (2013). On growth and formlets: Sparse multi-scale coding of planar shape. *Image and Vision Computing, 31*, 1–13.

El-Gaaly, T., Froyen, V., Elgammal, A. M., Feldman, J., & Singh, M. (2015). A Bayesian approach to perceptual 3D object-part decomposition using skeleton-based representations. *AAAI Conference on Artificial Intelligence*, 3762–3768.

Fahlman, B. D. (2011). *Materials chemistry*. Springer Science & Business Media.
Feldman, J. (1992). Constructing perceptual categories. In *Proceedings of the IEEE international conference on computer vision* (pp. 244–250).
Feldman, J. (1995) Formal constraints on cognitive interpretations of causal structure. In *Proceedings of the I.E.E.E. workshop on architectures for semiotic modeling and situation analysis, Monterey, CA*.
Feldman, J. (1997). The structure of perceptual categories. *Journal of Mathematical Psychology, 41*, 145–170.
Feldman, J., & Singh, M. (2005). Information along contours and object boundaries. *Psychological Review, 112*, 243–252.
Feldman, J., & Singh, M. (2006). Bayesian estimation of the shape skeleton. *Proceedings of the National Academy of Sciences, 103*, 18014–18019.
Feldman, J., Singh, M., Briscoe, E., Froyen, V., Kim, S., & Wilder, J. D. (2013). An integrated Bayesian approach to shape representation and perceptual organization. In S. Dickinson & Z. Pizlo (Eds.), *Shape perception in human and computer vision: An interdisciplinary perspective*. Springer Science & Business Media.
Ferziger, J. H., & Peric, M. (2012). *Computational methods for fluid dynamics*. Springer Science & Business Media.
Field, D. J., Hayes, A., & Hess, R. F. (1993). Contour integration by the human visual system: Evidence for a local "association field". *Vision Research, 33*, 173–193.
Fischer, B., & Modersitzki, J. (2008). Ill-posed medicine—An introduction to image registration. *Inverse Problems, 24*, 034008.
Fleming, R. W. (2015). Predicting shape variations from single exemplars. *Journal of Vision, 15*(12), 1126.
Froyen, V., Feldman, J., & Singh, M. (2015). Bayesian hierarchical grouping: Perceptual grouping as mixture estimation. *Psychological Review, 122*, 575–597.
Geisler, W., Perry, J., Super, B., & Gallogly, D. (2001). Edge co-occurrence in natural images predicts contour grouping performance. *Vision Research, 41*, 711–724.
Graf, M. (2006). Coordinate transformations in object recognition. *Psychological Bulletin, 132*, 920–945.
Gregory, E., & McCloskey, M. (2010). Mirror-image confusions: Implications for representation and processing of object orientation. *Cognition, 116*, 110–129.
Grossberg, S., & Mingolla, E. (1985). Neural dynamics of perceptual grouping: Textures, boundaries, and emergent segmentations. *Perception & Psychophysics, 38*, 141–171.
Haeckel, E. (1874). *Anthropogenie: Oder, Entwicklungsgeschichte des Menschen ["Anthropogeny: Or, the Evolutionary History of Man"]*. Leipzig: Engelmann.
Hahn, U., Chater, N., & Richardson, L. B. (2003). Similarity as transformation. *Cognition, 87*, 1–32.
Hahn, U., Close, J., & Graf, M. (2009). Transformation direction influences shape-similarity judgments. *Psychological Science, 20*, 447–454.
Hoffman, D. D., & Richards, W. A. (1984). Parts of recognition. *Cognition, 18*, 65–96.
Imai, S. (1977). Pattern similarity and cognitive transformations. *Acta Psychologica, 41*, 433–447.
Inostroza-Brito, K. E., Collin, E., Siton-Mendelson, O., Smith, K. H., Monge-Marcet, A., Ferreira, D. S., ... Mata, A. (2015). Co-assembly, spatiotemporal control and morphogenesis of a hybrid protein–peptide system. *Nature Chemistry, 7*, 897–904.
Kim, N. G., Effken, J. A., & Shaw, R. E. (1995). Perceiving persistence under change and over structure. *Ecological Psychology, 7*(3), 217–256.
Kimia, B. B., Tannenbaum, A. R., & Zucker, S. W. (1995). Shapes, shocks, and deformations I: The components of two-dimensional shape and the reaction-diffusion space. *International Journal of Computer Vision, 15*, 189–224.
Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3? Perception, 36th ECVP Abstract Supplement.
Koenderink, J. J., Kappers, A. M., Pollick, F. E., & Kawato, M. (1997). Correspondence in pictorial space. *Perception & Psychophysics, 59*, 813–827.
Koenderink, J. J., Van Doorn, A. J., Kappers, A. M., & Todd, J. T. (1997). The visual contour in depth. *Perception & Psychophysics, 59*, 828–838.
Koffka, K. (1935/1965). *Principles of gestalt psychology*. Harcourt, Brace & World.
Kourtzi, Z., & Shiffrar, M. (2001). Visual representation of malleable and rigid objects that deform as they rotate. *Journal of Experimental Psychology: Human Perception and Performance, 27*, 335–355.
Kubilius, J., Bracci, S., & Op de Beeck, H. P. (2016). Deep neural networks as a computational model for human shape sensitivity. *PLoS Computational Biology, 12*(4), e1004896.
Leyton, M. (1986a). A theory of information structure I. General principles. *Journal of Mathematical Psychology, 30*, 103–160.
Leyton, M. (1986b). A theory of information structure II. A theory of perceptual organization. *Journal of Mathematical Psychology, 30*, 257–305.
Leyton, M. (1989). Inferring causal history from shape. *Cognitive Science, 13*, 357–387.
Leyton, M. (1992). *Symmetry, Causality, Mind*. Cambridge, Mass: MIT Press.
Leyton, M. (2012). *Process grammar: The basis of morphology*. Springer.
Ma, J., Zhao, J., & Yuille, A. L. (2016). Non-rigid point set registration by preserving global and local structures. *IEEE Transactions on Image Processing, 25*, 53–64.
Mark, L. S., Shaw, R. E., & Pittenger, J. B. (1988). Natural constraints, scales of analysis, and information for the perception of growing faces. In T. R. Alley (Ed.), *Social and applied aspects of perceiving faces: Resources for ecological psychology* (pp. 11–49). Hillsdale, NJ: Lawrence Erlbaum Associates.
Mark, L. S., & Todd, J. T. (1985). Describing perceptual information about human growth in terms of geometric invariants. *Perception & Psychophysics, 37*, 249–256.
Medendorp, W. P. (2011). Spatial constancy mechanisms in motor control. *Philosophical Transactions of the Royal Society B: Biological Sciences, 366*, 476–491.
Moran, S., & Leiser, D. (2002). The limits of shape constancy: Point-to-point mapping of perspective projections of flat figures. *Behaviour & Information Technology, 21*, 97–104.
Movahedi, V., & Elder, J. H. (2010). Design and perceptual validation of performance measures for salient object segmentation. In *Proceedings of the 7th IEEE computer society workshop on perceptual organization in computer vision (POCV)* (pp. 49–56).
Myronenko, A., & Song, X. (2010). Point set registration: Coherent point drift. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 32*, 2262–2275.

Norman, J. F., Phillips, F., & Ross, H. E. (2001). Information concentration along the boundary contours of naturally shaped solid objects. *Perception, 30*, 1285–1294.
Oliveira, F. P., & Tavares, J. M. R. (2014). Medical image registration: A review. *Computer Methods in Biomechanics and Biomedical Engineering, 17*(2), 73–93.
Ons, B., & Wagemans, J. (2011). Development of differential sensitivity for shape changes resulting from linear and nonlinear planar transformations. *i-Perception, 2*(2), 121.
Ons, B., & Wagemans, J. (2012). Generalization of visual shapes by flexible and simple rules. *Seeing and Perceiving, 25*, 237–261.
Paluch, E., & Heisenberg, C. P. (2009). Biology and physics of cell shape changes in development. *Current Biology, 19*, 790–799.
Panis, S., Vangeneugden, J., & Wagemans, J. (2008). Similarity, typicality, and category-level matching of morphed outlines of everyday objects. *Perception, 37*, 1822–1849.
Paulun, V. C., Kawabe, T., Nishida, S. Y., & Fleming, R. W. (2015). Seeing liquids from static snapshots. *Vision Research, 115*, 163–174.
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*, 1481–1492.
Phillips, F., Todd, J. T., Koenderink, J. J., & Kappers, A. M. (2003). Perceptual representation of visible surfaces. *Perception & Psychophysics, 65*, 747–762.
Pinna, B. (2010). New Gestalt principles of perceptual organization: An extension from grouping to shape and meaning. *Gestalt Theory, 32*, 11–78.
Pittenger, J. B., & Shaw, R. E. (1975). Aging faces as viscal-elastic events: Implications for a theory of nonrigid shape perception. *Journal of Experimental Psychology: Human Perception and Performance, 1*, 374–382.
Pittenger, J. B., & Todd, J. T. (1983). Perception of growth from changes in body proportions. *Journal of Experimental Psychology: Human Perception and Performance, 9*, 945–954.
Richards, W., Feldman, J., & Jepson, A. (1992). From features to perceptual categories. *Proceedings of the British Machine Vision Conference*, 99–108.
Rosengren, K. S., Gelman, S. A., Kalish, C. W., & McCormick, M. (1991). As time goes by: Children's early understanding of growth in animals. *Child Development, 62*, 1302–1320.
Schmidt, F., Spröte, P., & Fleming, R. W. (2016). Perception of shape and space across rigid transformations. *Vision Research, 126*, 318–329.
Scholl, B. J., & Tremoulet, P. D. (2000). Perceptual causality and animacy. *Trends in Cognitive Sciences, 4*(8), 299–309.
Shaw, R., & Pittenger, J. (1977). Perceiving the face of change in changing faces: Implications for a theory of object perception. In R. Shaw & J. Bransford (Eds.), *Perceiving, acting, and knowing: Toward an ecological psychology* (pp. 103–132). Hillsdale, NJ: Erlbaum Associates.
Shepard, R., & Cooper, L. (1982). *Mental images and their transformations*. Cambridge, MA: MIT Press.
Spröte, P., & Fleming, R. W. (2013). Concavities, negative parts, and the perception that shapes are complete. *Journal of Vision, 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, 126*, 330–346.
Thompson, D. W. (1942). *On growth and form*. Cambridge, UK: Cambridge University Press.
Todd, J. T. (1982). Visual information about rigid and nonrigid motion: A geometric analysis. *Journal of Experimental Psychology: Human Perception and Performance, 8*, 238–252.
Todd, J. T. (1984). The perception of three-dimensional structure from rigid and nonrigid motion. *Perception & Psychophysics, 36*(2), 97–103.
Todd, J. T., Weismantel, E., & Kallie, C. S. (2014). On the relative detectability of configural properties. *Journal of Vision, 14*(1), 18.
Vrins, S. A., de Wit, T. C., & van Lier, R. J. (2009). Bricks, butter, and slices of cucumber: Investigating semantic influences in amodal completion. *Perception, 38*, 17–29.
Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., & von der Heydt, R. (2012). A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure–ground organization. *Psychological Bulletin, 138*, 1172–1217.