Acta Psychologica 128 (2008) 331–338

Contents lists available at ScienceDirect

Acta Psychologica

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

Visual judgment of similarity across shape transformations: Evidence for a compositional model of articulated objects

Elan Barenholtz [a,*], Michael J. Tarr [b]

[a] Department of Psychology, Florida Atlantic University, 777 Glades Road, Boca Raton, FL 33431, United States
[b] Department of Cognitive and Linguistic Sciences, Brown University, Providence, RI, United States

ARTICLE INFO

Article history:
Received 1 November 2007
Received in revised form 10 March 2008
Accepted 11 March 2008
Available online 7 May 2008

PsycINFO classification:
2323

Keywords:
Visual perception
Shape perception
Visual recognition
Invariance
Articulation
Biological motion

ABSTRACT

A single biological object, such as a hand, can assume multiple, very different shapes, due to the articulation of its parts. Yet we are able to recognize all of these shapes as examples of the same object. How is this invariance to pose achieved? Here, we present evidence that the visual system maintains a model of object transformation that is based on rigid, convex parts articulating at extrema of negative curvature, i.e., part boundaries. We compared similarity judgments in a task in which subjects had to decide which of the two transformed versions of a 'base' shape—one a 'biologically valid' articulation and one a geometrically similar but 'biologically invalid' articulation—was more similar to the base shape. Two types of comparisons were made: in the *figure/ground-reversal*, the invalid articulation consisted of exactly the same contour transformation as the valid one with reversed figural polarity. In the *axis-of-rotation reversal*, the valid articulation consisted of a part rotated around its concave part boundaries, while the invalid articulation consisted of the same part rotated around the endpoints on the opposite side of the part. In two separate 2AFC similarity experiments—one in which the base and transformed shapes were presented simultaneously and one in which they were presented sequentially—subjects were more likely to match the base shape to a transform when it corresponded to a legitimate articulation. These results suggest that the visual system maintains expectations about the way objects will transform, based on their static geometry.

1. Introduction

1.1. Pose invariance

Many objects, particularly biological ones, can assume multiple, very different shapes due to natural movements, such as the articulation of body parts and limbs. For example, a human hand can take on an essentially infinite number of poses, due to articulations of its fingers. Yet, our visual system can effortlessly identify these different articulated shapes—even ones we have never seen before—as hands. This ability to classify radically different images as representing a single object type or class is a form of visual invariance; that is, invariance to pose under articulations. However, unlike other sources of image variability, such as those due to position, scale and orientation, all which are produced by changes in the spatial relationship between the observer and the object, the variation due to pose reflects transformations of the object itself; thus, recognition methods that are based on holistic image normalization (e.g., Huttenlocher & Ullman, 1990, 'method of alignment') are not feasible. In addition, we are able to recognize *novel* poses of complex articulated objects that can assume a wide variety of possible shapes. This ability to generalize from one pose of an object to new poses seems to be based on learning from a limited number of examples, that is, for pose invariance we do not seem to rely on multiple 'views' independent of generalization mechanisms. This point can be illustrated by observing that our propensity for extrapolating new poses is not limited to familiar objects. For example, if asked to name the blobby shape in Fig. 1, you presumably could not do so. However, if asked to predict how this object might articulate, you probably have strong intuitions. That is, we seem to maintain some prior expectations with regard to the way shapes typically transform.

1.2. Part-based representations

What principles might the visual system apply to extrapolate the possible shapes of an articulating object based on a small number of examples? We suggest that this form of inference depends on a *structural* representation that includes some set of *parts* and

[*] Corresponding author. Tel.: +1 (561)297 3433; fax: +1 (561)297 2160.
E-mail address: elan.barenholtz@fau.edu (E. Barenholtz).

0001-6918/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.actpsy.2008.03.007

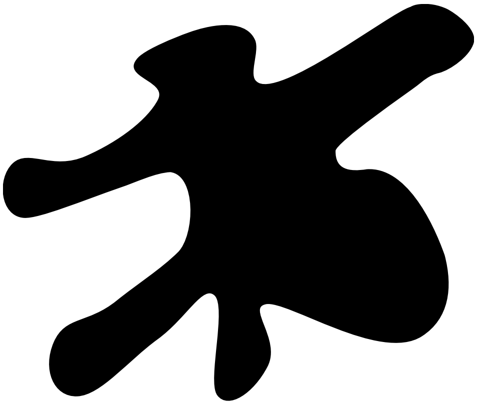

Fig. 1. A novel shape with an intuitive articulatory structure.

the *relations* between them (Barenholtz & Tarr, 2007; Biederman, 1987; Hummel & Stankiewicz, 1997; Marr & Nishihara, 1978).[1] For example, in the illustration above, your expectations probably involved the motion of 'arm-like' projections, a formulation that implicitly assumes a division of the objects into parts. In addition, you might have anticipated that the motion of these parts will be constrained to rotations hinged at the 'body' to which they are connected. Presumably, these intuitions are based on the natural motion of biological organisms: the contour shape of vertebrates is constrained by layers of fat and muscle tissue over a skeleton of rigid, connected bones; biological articulations (i.e., 'biological motion') arise when the underlying rigid bones rotate at their connecting joints.

The idea that complex objects may be represented on the basis of constituent parts is well represented in the psychological literature. A number of influential theories of recognition hold that objects are indexed on the basis of a set of canonical parts and the relations between them (Biederman, 1987; Hummel & Stankiewicz, 1997; Marr & Nishihara, 1978). Perhaps the most prominent example of a structural theory is Biederman's recognition by components (RBC), which holds that viewed objects are matched to their stored representations on the basis of a set of geometrically simple canonical units ('geons', such as cones and bricks) and their interrelations. A major theoretical advantage of RBC is that it provides an economical method for achieving viewpoint-invariance, since a limited number of individual geons may be identified on the basis of properties that are stable under rotations in depth.

In addition to theories that depend on a set of predetermined parts, there is another line of work in the psychological literature that assumes that the visual system extracts parts '*ad-hoc':* on the basis of geometric properties applied directly to the viewed shape rather than by matching features to a predetermined lexicon. Unlike RBC theory, partitioning based on local object properties can lead to arbitrarily shaped parts. An important advance in this regard was Hoffman and Richards' 'minima rule' (1984), which states that objects are divided at regions of maximal negative curvature. There is substantial direct empirical evidence that minima play an important role in object parsing (Barenholtz & Feldman, 2003; DeWinter & Wagemans, 2006). Surprisingly, much less work has addressed the potential function of such segmentation. One possibility, of course, is that partitioning simply serves as a front-end of the kind of lexical system mentioned above—parts are extracted in order to match them to a stored lexicon of generic features (e.g. geons). An alternative view, however—one that we explore in the current paper—is that geometric partitioning might be intended to determine a set of features *specific* to a particular class or objects, that is, not generic to all objects, in order to generalize across instances of the same object or class, under varying articulatory poses.

1.3. Part-based articulations

Biological articulations arise when globally convex[2] body parts (limbs, or parts of limbs) change their relative angle via rotations of the underlying skeletal joints. It has long been known that the visual system is extremely sensitive to the regularities of motion produced by joint-based articulations. In 'biological motion', observers have been shown to recognize numerous characteristics of motion on the basis of minimal point-light displays (Johansson, 1973). Typical biological motion displays involve points sampled from the interior of the shape (usually at the joints themselves), rather than along the outer contour. However, the articulation of limbs produces signature deformations of the bounding contour as well (Barenholtz & Feldman, 2006). We note that such articulations carry the following geometric regularities with regard to the contour shape (see Fig. 2):

(1) *Preservation of part-shape*: While the global shape may change, the shape of individual articulating parts, defined by parsing at the concave minima (see above), will remain rigid.

(2) *Preservation of part-boundary location*: Under articulations, the parts will remain connected to one another at the same locations within the part—that is, at the part boundaries.

It should be noted that these regularities are an ideal: actual biological articulations typically involve complex deformations at the boundary due to loose fitting skin, tissue compression, etc. and neither shape nor boundary location is preserved exactly. However, the deformations of parts and their boundaries are relatively minimal compared with the global deformation of the object due to the rotation of the part and the regularities described are approximately true of articulations.

Knowledge of these geometric regularities may provide the visual system with a capacity to 'predict' the kinds of shape changes an object may undergo. In a recent study, Barenholtz and Feldman (2006) found evidence that the visual system employs the geometric regularities of articulation when interpreting dynamic contour deformations. In these experiments, subjects performing a figure/ground assignment task on moving displays showed a strong bias towards assigning figure so that motion was consistent with convex parts articulating at concave minima[3]; that is, they showed a preference for contour motion consistent with the geometry of biological articulations (see Fig. 3).

[1] The term 'structure' has often been used as shorthand to denote the nature of visual *features*. In particular, "structural" approaches are usually meant to imply 3D shape-models of features, such as Geons (Biederman, 1987), while 'view-based' theories have often been associated with 2D images as features (Edelman, 1993). Indeed, the second author of this paper has often been associated with this latter approach. However, the issue of the particular nature of discrete features (e.g., Geons or 2D-patches) is orthogonal to the sense of structure intended here: the separate representation of features and the *relations* between them, independent of the particular nature of the features (for more detail on our perspective on this issue, see Barenholtz & Tarr, 2007).

[2] While biological limbs may contain concavities, the *global* shape of such parts —as defined by partitioning the part at its boundaries with the rest of the object—is convex. For example, a human arm contains concavities (at the elbow and between the fingers). However, we would say that it is 'globally' convex since it protrudes out of, not into, the rest of the body.

[3] While articulations may sometime also involve a convex fulcrum in addition to a concave fulcrum (for example, an elbow), they *always* involve at least one concave fulcrum, which forms the part boundary.

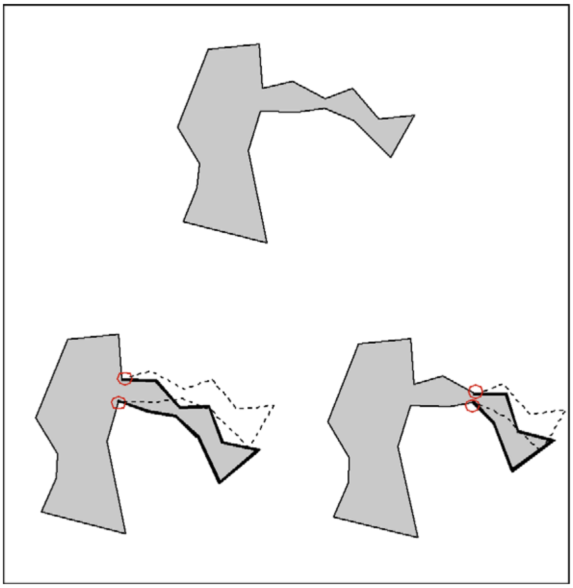

Fig. 2. The two shapes on the bottom are possible articulations of the top shape. The articulated part, defined by the section of the contour from the concave part boundaries (circled) outward, maintains its shape. The part boundaries remain static under articulations and maintain their position.

1.4. The current study

In the current paper, we explore the idea that the same regularities which allow the visual system to perceive articulatory motion may also provide the leverage to 'infer' such motion among static views of objects, that is, in order to recognize the same object under different articulatory poses. According to this view, two shapes will be treated as corresponding to the same object if and only if they share parts—defined by parsing at negative minima—that are connected in the same locations—the negative minima themselves. We explored this hypothesis by asking the following experimental question: are two shapes that are biologically 'valid' articulations of one another, according to the constraints described above, psychologically more *similar* to one another? That is, are observers more likely to perceive the two shapes as corresponding to the same 'object' as compared to other shapes that correspond to biologically 'invalid' deformations that are geometrically comparable? We employed a simple experimental design in which subjects judged which of the two target shapes was more similar to a single 'base' shape. Each of the two target shapes was a transformed version of the base shape; however, only one of the two targets corresponded to a valid articulation of the base shape—as defined by the constraints described above—while the other target (the 'distractor') did not. A critical challenge in testing this hypothesis is to isolate, to the greatest extent possible, the critical variables that determine whether a transformation is a valid articulation or not. In the present study we used two different methods for establishing a basis of comparison to valid articulations, each with particular advantages: (1) figure/ground-reversal; (2) axis-of-rotation reversal.

1.4.1. Figure/ground-reversal

As demonstrated by Attneave (1971), who showed that an egg-shape divided in two by a single contour produces two very different looking halves, the assignment of figure and ground to a contour determines its perceived shape. Since figural assignment determines sign of curvature—and articulation depends critically on the sign of curvature of the deforming contour—the reversal of figure and ground also determines whether a given contour deformation is a valid articulation or not. This allowed us to create stimuli in which the *exact same contour deformation* was presented twice: once as a valid articulation of a part rotating at concavities and once where this interpretation did not apply. Fig. 4 demonstrates the basic idea: The pivoting of a single 'part' of a contour (top row) may correspond to a valid articulation or not depending on the figural assignment to the contour. In the middle row, the figural assignment (to the left) means that the pivoting part is convex, articulating at concavities—a biologically 'valid' articulation; in the bottom row the 'part' is concave, articulating at convexities—a biologically 'invalid' articulation. It is important to note that the geometric transformation is *locally* identical in all of these cases; only the figural assignment to the contours is different. However, only the valid articulations preserve part-shape. In our experiment, we generated a set of these 'articulated' contours, referred to as 'base and transformed' contours and used them to divide a polygonal shape into two closed shapes with a shared contour with opposite figural polarity (see Fig. 6 for details of stimulus construction). Then we tested to see whether there was a tendency to see the transformed shape in which the figural assignment was consistent with a valid articulation (based on the figural assignment) compared with a transformation that was not. An important detail of the experimental design was that each dividing contour actually had two transformed versions (see Methods), each of which was a valid articulation under one figural assignment and not the other. This

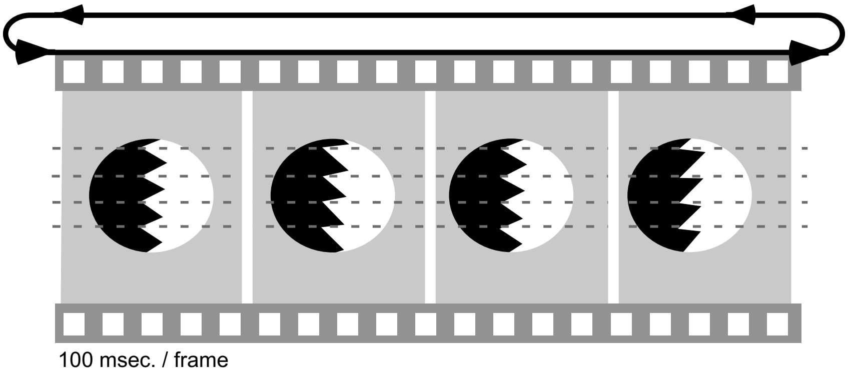

Fig. 3. An illustration of a dynamic stimulus used by Barenholtz and Feldman (2006). In this example, all the vertices on one side (right in this example) of a jagged 'comb' were shifted up and down in a repeated 4-frame cycle. Subjects showed a strong preference for a figural assignment in which the moving vertices were convex and the stationary vertices were concave; that is, the black "teeth" appear to wave up and down in front of a white background, consistent with convex parts, articulating at concave "joints." An example of these stimuli can be viewed at: http://psy.fau.edu/~barenholtz/Demos/articulation_fg_stim.gif.

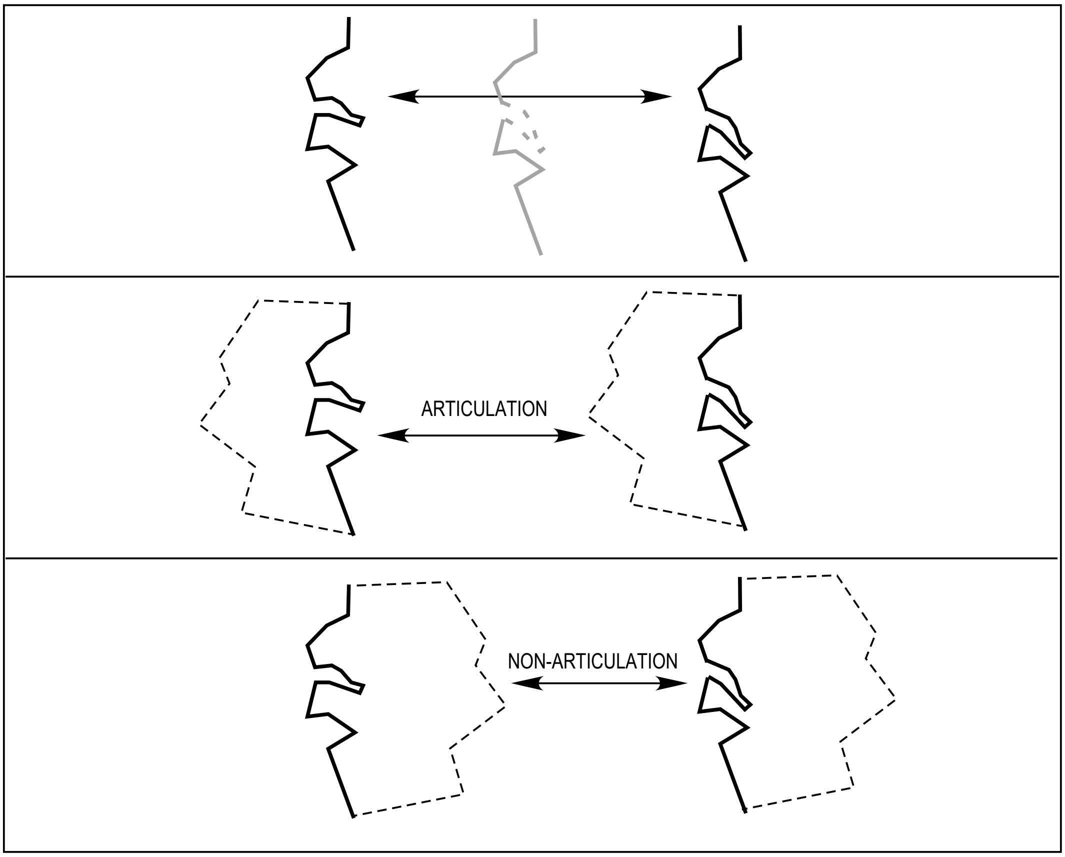

Fig. 4. Figure–ground-reversal. The exact same contour transformation can be a biologically 'valid' or 'invalid' articulation, depending on the figure/ground assignment. *Top*: A part of a contour is rotated at curvature extrema. *Middle*: When figure is completed to the left (in this example), the part is convex, rotating at concavities—a valid articulation. *Bottom*: When figure is completed to the right, the part is concave, rotating at convexities—an invalid articulation.

allowed us to see whether preferences flipped (with figural assignment) for the *identical contour comparison* across separate trials. If there is a preference for the deformation only in the case where, because of figure/ground assignment, it corresponds to a valid articulation, then this points to a preference for articulations *per se*, as compared with identical contour deformations that are not articulations.

1.4.2. Axis-of-rotation reversal

The figure/ground-reversal paradigm used here has the desirable characteristic of equating geometry across the two comparison transformations; the only difference between them is their sign of curvature. Another way of characterizing the difference between the two transformations—the biologically valid articulation and invalid articulation—is that one figural assignment preserves part-shape (defined by parsing at negative minima) while the other figural assignment does not. Thus, a preference for a contour deformation under the figural assignment that corresponds to a valid articulation (i.e., the rotation of a convex part at concavities) would provide support for Regularity 1, preservation of part-shape.

However, this does not directly test Regularity 2, which requires that the spatial relations between parts must be constrained so that part boundaries do not change location. This is particularly important in establishing that invariance to pose requires explicit encoding of spatial relations independent of the features. To test Regularity 2, we generated a different set of shapes in which two different transformations preserved part-shape and orientation equally, but where only one preserved the location of the part boundaries. In this case both transformations consisted of a shear of a part, which is perceptually similar to a rotation (Fig. 5; also see Methods section for details of stimulus construction). In one

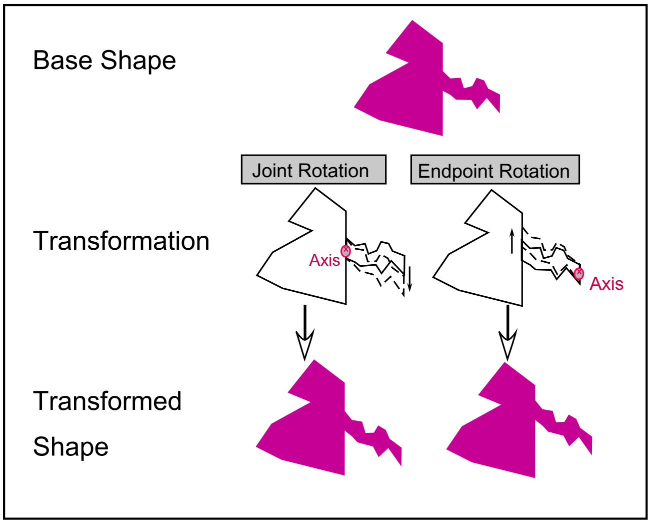

Fig. 5. The 'axis-of-rotation reversal'. See text for details of construction. A shearing of a part is a biologically valid articulation only when it preserves the part-boundary locations. Note that the two transformed parts resulting from the shearing are virtually identical in shape; only their position relative to the rest of the shape is different.

case, the 'rotation' was anchored at the part boundaries ('joint rotation'). In the other the rotation was anchored at the opposite endpoint of the part ('endpoint rotation'). In both instances, part-shape and orientation are preserved equally and the orientation of the two transformed parts is identical. However, the location of the part boundaries is only preserved in the joint rotation. Thus, any tendency to favor one transformation over the other cannot be due to the preservation of parts or their orientations; instead, it points to a preference for transformations that observe specific *relational* constraints.

Subjects performed a two-alternative forced choice task in which they decided which of the two transformed shapes looked most similar to a single base shape (note that, unlike typical 2AFC tasks, there was no 'correct' answer). Two separate experiments, using different subjects, were performed using the same set of stimuli. In Experiment 1, the simultaneous experiment, the base-shape and the two transformed shapes were presented simultaneously until the subject responded. In Experiment 2, the sequential experiment, the base shape and transformed shapes were presented sequentially separated by a brief inter-stimulus interval. The sequential experiment was performed in order to minimize the role of deliberate, point-by-point comparison between the probe and base shapes, instead, forcing subjects to rely on their memory of the shape. The method used in the sequential experiment is similar to the same-different sequential matching task employed in numerous studies of visual recognition, in which subjects must determine whether two items presented in sequence are identical (see, for example Le Grand, Mondloch, Maurer, & Brent, 2001; Riesenhuber, Jarudi, Gilad, & Sinha, 2004; Yovel & Duchaine, 2006). This methodology is particularly useful when, like in the current experiment, the stimulus differences are subtle and thus, unlikely to be captured by long-term memory tasks. However, note that, unlike these other experiments, in the current study *neither* target was identical to the original stimulus. Indeed, we make no predictions as to whether legitimate articulations are more or less likely to be *confused* with an original shape as compared with other shape changes in a short-term memory task; articulations may be highly salient and easily detected. However, we believe that using a memory task such as sequential presentation provides a closer approximation of typical recognition than a simultaneous comparison, which can be accomplished using only visible perceptual information.

2. Methods

2.1. Subjects

Sixteen Brown University undergraduates—8 in the simultaneous experiment and a different group of 8 in the sequential experiment—participated in the study in exchange for payment.

2.2. Stimuli

The stimulus set was identical in both experiments and consisted of computer generated complex (i.e., multi-part) polygonal shapes designed using the Canvas drawing program, and measuring approximately 2.3° visual angle in height and 2.6° in width, presented on an Apple iMac G5 20" LCD screen. All the stimuli used in the experiment can be seen online at: http://psy.fau.edu/~barenholtz/articulation_stimuli.html. The stimulus set consisted of two shape sets, corresponding to two different comparison types: Figure/ground-reversal and axis-of-rotation reversal, described below. The *figure/ground-reversal* set consisted of 20 groups of six shapes—two 'base' shapes and four transformed shapes—for a total of 120 stimulus shapes. Each group was constructed using a triad of polygonal contours (the 'dividing contours') consisting of a 'base' contour and two transformations: a 'righty' part and 'lefty' part rotation (Fig. 6). The part rotations were generated by choosing two 'anchor points' along the contour, which were vertices with equal curvature polarity (i.e., they were both either concave or convex vertices depending on figure–ground assignment). These two points enclosed a region of the contour (a 'part'), which, depending on the particular points, was projecting from the contour on either the right or the left ('righty' and 'lefty' parts in the figure). Each part was then rotated between 10° and 25° around an axis defined as the midpoint between the two anchor vertices, and reconnecting the part by extending its edge. Each of the triad of contours was then used to divide the same, bilaterally symmetrical shape into two separate shapes for a total of six shapes: two base shapes and four transformed shapes. Note that each righty and lefty contour forms a valid and invalid articulation of the base contour depending on the figural assignment.

The *axis-of-rotation reversal* set consisted of 20 groups of 3 shapes—consisting of a base shape and two transformed shapes—for a total of 60 shapes. Each shape was a closed polygon with several extending 'parts' defined by partitioning at negative minima

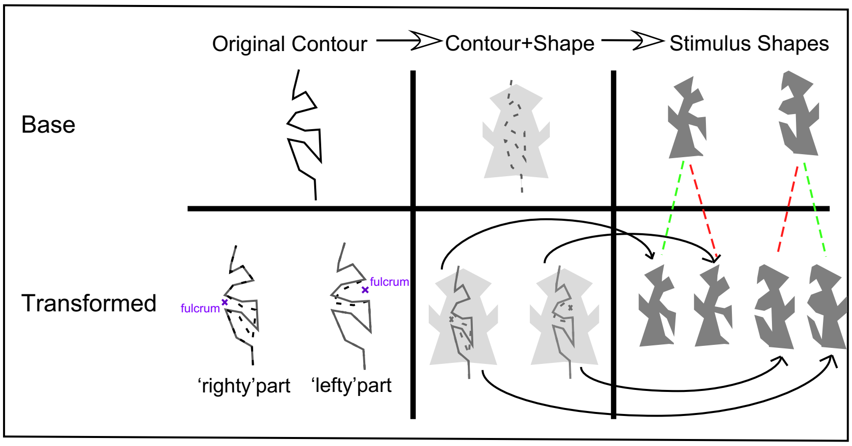

Fig. 6. Construction of the figure–ground-reversal stimuli. See text for details. *Left:* A single base contour (top) is used to generate two 'transformed' contours (bottom), with 'righty' and 'lefty' part rotations. *Middle*: Each contour (the base and two transformed) are used to divide the same shape in two. *Right:* The resulting stimulus shapes. Each dividing contour—the base, lefty and righty—produces two stimulus shapes—two base and four transformed, for a total of six shapes. Note that each base shape has a 'valid' (green dashed arrows) and 'invalid' (red dashed arrows) transform match, which is the reverse of the other base shape. These triads—base, valid and invalid—were the basis of comparison in the experiment. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(see Fig. 5 for further details of the shape construction). There were two types of transformed shapes for each base shape: a 'joint rotation' and an 'endpoint rotation'. Both rotation types were formed by shearing a polygonal part around an axis[4]; in the joint rotation, this was the midpoint between the concave part boundaries; in the endpoint rotation it was the midpoint between the two points farthest from the part boundary (i.e., the ends of the part). The region in the area of the part boundary was always a straight edge. This ensured that no discernible shape landmarks were altered as the part translated during the endpoint rotation. In the current experiment, the two transformations were shears of equal magnitude, differing in their axes-of-rotation. As can be seen in Fig. 5, the parts produced by these two shear transformations are nearly identical in terms of subjective shape. Perhaps more importantly, the two transformations preserve *true* shape approximately equally across the two conditions since neither axis-of-rotation preserves local shape any better than the other.

2.3. Procedure

On each trial, the observer was presented with a centered fixation cross for 500 ms. In the simultaneous experiment, this was then replaced by a display containing the base shape on the top half of the screen and two probes on the lower left and right sides of the screen. In the sequential experiment, the fixation cross was replaced by the base stimulus, presented in the center of the screen for 1 s, followed by a white screen for 2 s, and then the probes to the right and left of the center. The subject's task was to choose which of the probes (left or right) was more similar to the base shape. On figure/ground-reversal trials, the targets consisted of the 'righty' and 'lefty' transformations of the base shape, On the axis-of-rotation trials, they consisted of the two transformation types (endpoint and joint rotations). On each trial, one of the two comparison types was represented. There were 60 trials consisting of 40 figure–ground-reversal and 20 axis-of-rotation comparisons. The two types of comparison types were randomly distributed throughout the single experimental block performed by each subject.

3. Results

Fig. 7 shows the percentage of trials on which subjects chose a biologically valid articulation as being more similar than the biologically invalid articulation, across the two reversal types (figure/ground and axis-of-rotation) for Experiments 1 and 2.

3.1. Experiment 1 – simultaneous

Collapsing across both reversal types, there was a preference to choose the valid articulations over the non-articulations on 70% of trials (SE = 2%). A single-sample *t*-test (testing for a difference between the observed mean and a hypothetical population with no bias) found that this preference was significant ($t(7) = 7.534$, $p = .0003$). Distinguishing by reversal type, for the figure/ground-reversal there was a preference for valid articulations, that is, the transformed shape consistent with a convex part rotating at concavities, (mean = 66%, SE = 5%). This preference was significant ($t(7) = 3.494$, $p = .0101$). For the axis-of-rotation reversal there was a preference for valid articulations; that is, the transformed shape consistent with a rotation at the part boundaries, (mean = 75%, SE = 5%). This preference was significant ($t(7) = 4.448$, $p = .0043$).

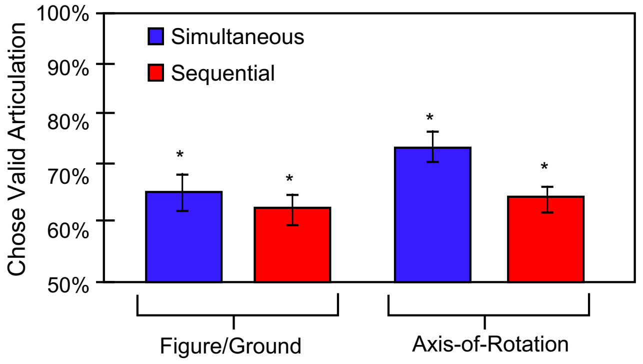

Fig. 7. Percentage of trials on which respondents chose the target shape consistent with a biologically valid articulation for the two experiments, simultaneous and sequential simpler features (i.e., parts) and the relations between them. Asterisks denote a significant preference for the 'valid' articulation.

3.2. Experiment 2 – sequential

Across both comparison types, there was a preference to choose the valid articulations over the non-articulations (mean = 62%, SE = 2.5%).This preference was significant: ($t(9) = 4.730$, $p = .0011$). Distinguishing by comparison type, for the figure/ground-reversal there was a preference for valid articulations (mean = 60%, SE = 2%). This preference was significant $t(9) = 4.730$, $p = .0011$. For the axis-of-rotation reversal there was a preference for valid articulations (mean = 61%, SE = 3%). This difference was significant ($t(9) = 3.608$, $p = .0057$).

4. Discussion

The results of these two experiments suggest that observers view transformations of shape that carry a biologically valid interpretation (i.e., an articulation) as leading to greater similarity than geometrically similar transformations that do not carry such an interpretation. This preference for valid articulations, while significant, were not always strong, especially in the sequential condition of the figure/ground-reversal conditions. Why, if valid articulations are viewed as the 'same' object, should they not be preferred in virtually all of the cases? First, it should be noted that the differences between the different transformations in the figure/ground-reversal condition were subtle and might have been difficult to detect with only a second of preview time. However, it should also be noted that in the figure/ground-reversal condition, while the invalid articulations did not preserve part-shape, they did for the most part preserve part *orientation* with higher fidelity than the valid articulations, where the orientation of the articulated part changed dramatically. Indeed, as mentioned earlier, it is possible that the differences due to valid articulations in this experiment are equally (or even more) visually *salient* than invalid ones, a possibility that should be addressed by future research. Thus, we do not interpret our results to mean that subjects did not notice the changes due to articulations; rather subjects noticed both kinds of transformation but maintained a preference for the preservation of part-shape over part-orientation on most trials because the former corresponds to natural articulations.

Our current results provide evidence that a part-based model may account for some of our ability to achieve 'pose invariance' to changes in shape due to natural articulations. This model is con-

[4] A shear is an affine transformation which preserves certain important shape properties such as parallelism and colinearity. Human subjects show a strong capacity for determining equivalence among affine transformations of patterns (Wagemans, van Gool, Lamote, & Foster, 2000). As applied to our stimuli, the shear looks very similar to a rigid rotation. The shear was used, rather than a true rotation, because it allowed us to approximate a rotation (to which it appears perceptually similar) while preserving the location of the connections between the parts.

sistent with more general part-based approaches to visual recognition. In particular, these results suggest a potential utility for '*ad-hoc*' partitioning schemes such as the 'minima rule' (Hoffman & Richards, 1984) that do not carry the efficiency of a predetermined part lexicon: one reason to divide an object into parts is because the shape of the parts and aspects of their spatial relations remain constant across articulations. While the current results apply most obviously to animate biological that articulate dynamically, the same model of 'articulation' might apply to inanimate classes of objects as well. To begin with, some inanimate objects *do* articulate in a manner that is similar to biological objects, for example the folding temples of eyeglasses. As documented by DeWinter and Wagemans (2006), such structures are treated as parts, much like the limbs of a biological object. However, the current results may have application even to *rigid*, non-articulating objects; for certain classes of objects the variability that must be overcome is not because a single object can change its shape but because, across the entire class of objects, different 'poses' may be present. For example, while a four-legged chair's legs do not articulate, the orientation of the legs relative to the seat varies across examples of chairs (e.g., they may be perpendicular to the seat or at an angle). However, they do not vary arbitrarily: the legs of a chair are usually connected at around the same location of the seat, with a shallow range of angles that might be approximated by biological articulations. Thus, it is possible that the regularities described here with regard to biological objects may also be applicable to some classes of inanimate objects as well.

Of course, even with regard to biological objects, our current results do not provide details for a full-fledged model for achieving invariance to pose under 'real-world' conditions. Unlike our 2-D stimuli, the apparent shape (i.e., the retinal image) of the parts of 3-D objects changes with rotations in depth (which are likely to occur under typical articulations). The question of how the visual system overcomes such variation remains a critical—and controversial—problem in vision science, with an ongoing debate as to whether recognition relies on 'view-dependent' features (e.g., Tarr & Bülthoff, 1995), 'view-independent features' (e.g., Biederman, 1987), or both (Vanrie, Willems, & Wagemans, 2001). However, we argue that the potential solution to this more general problem is beyond the scope of the current discussion. Virtually any strategy for overcoming rotations in depth could be applied to segmented parts of objects, and the principles discussed here will be equally valid.

Thus, while the current results do not have any bearing on the question of viewpoint dependency, they do have implications for the general idea of 'structure' in models of vision (Footnote 1). In particular, these findings point to a role not just for parts, but the explicit encoding of *relations* between parts as well. As mentioned earlier, a number of influential theories of recognition have relied on models in which relational information is explicitly encoded (Biederman, 1987; Marr & Nishihara, 1978). From an empirical standpoint, a number of studies have looked explicitly at the role of relations in the encoding of complex objects (Hummel & Stankiewicz, 1997; Saiki & Hummel, 1998). However, a critical challenge in considering empirical evidence for structure is that it is difficult to rigorously distinguish between the encoding of features and their configurations. This is due to the fact that any particular configuration of features can—in theory—be represented in terms of a more 'complex', larger feature, composed of the smaller features in that particular configuration (for an extensive discussion of this issue see Barenholtz & Tarr, 2007). For example, a body of work concerning facial recognition considers the perception of artificial faces constructed to differ on the basis of their 'features' (i.e., eyes, nose, etc.) or their 'configuration' (i.e., spacing between features; Tanaka & Farah, 1993). However, a potential red herring in such work is that 'configural' differences may actually be encoded in terms of larger unitary features; for example, rather than encoding the distance between eyes, identification may be based on a single image fragment that incorporates both eyes (e.g., Zhang & Cottrell, 2005). Similarly, a number of recent studies have purported to show differential processing of parts and relations in a search task (Arguin & Saumier, 2004) and in a change-detection task (Keane, Hayward, & Burke, 2003). Both of these studies compared visual processing for objects that had shared parts but in different spatial relations vs. objects that had different parts but in a similar configuration; the underlying assumption of these studies is that the latter case—objects with different parts in similar configurations—are similar with regard to their *relational* information. However, we would argue that it is possible that the similarity between these objects is captured by considering their global form or image (which is usually more similar among objects that share configurations than those that share parts), without explicitly encoding the relations between separable parts (Barenholtz & Tarr, 2007).

Little experimental research has been performed that expressly addresses this ambiguity. While our present results likewise do not directly address this issue, we believe that the examples of articulation presented here provide evidence for structural representations in vision. First, Regularity 1 (preservation of parts), demonstrated in the figure/ground-reversal comparison, supports the idea of a parts-based approach. Importantly, because the geometric deformations were *identical* (independent of figural assignment) in both cases of the figure–ground-reversal, neither one can be said to preserve any global shape property better than the other. Rather, the most plausible explanation depends on the fact that only valid articulations preserve part-shape. Second, Regularity 2 (preservation of part-boundary location) demonstrated in the axis-of-rotation reversal shows that representing these parts alone, in the absence of relations, is insufficient because neither valid nor invalid articulations preserve any plausible parts with greater fidelity. What *is* preserved is the connectivity at the original part boundaries, a property which (unlike some of the examples in the studies mentioned above) cannot be captured by simply comparing 'global' shape; instead, it requires encoding explicit *relational /configural* information. In short, both features *and* their relations—that is compositional structure—are required to achieve invariance to articulatory pose.

Within both the computational and psychophysical literatures, the recent trend has been a move away from structural accounts to more image-driven theories of recognition using simple features of the image —usually patches (Riesenhuber & Poggio, 1999; Ullman, Vidal-Naquet, & Sali, 2002; Wallis & Rolls, 1997;). According to these theories the relations between individual features are not explicitly encoded; only the presence of a feature is considered as evidence (although, again, relations may be implicitly captured by *larger* features that encompass two relevant features). Our results on articulation suggest that, at a minimum, simple feature-based models of similarity, particularly those that rely on only relatively local features, will not be able to account for certain aspects of visual cognition; in particular, the kind of robust generalization characterized by visual invariance to pose appears to depend on some notion of separable parts or features *as well as* the relations between them. It is likely that other aspects of visual recognition do as well.

References

Arguin, M., & Saumier, D. (2004). Independent processing of parts and their spatial organization in complex objects. *Psychological Science, 15*, 629–633.

Attneave, F. (1971). Multistability in perception. *Scientific American, 225*, 62–71.

Barenholtz, E., & Feldman, J. (2003). Visual comparisons within and between object-parts: evidence for a single- part superiority effect. *Vision Research, 43*, 1655–1666.

Barenholtz, E., & Feldman, J. (2006). Determination of visual figure and ground in dynamically deforming shapes. *Cognition, 101*, 530–544.

Barenholtz, E., & Tarr, M. J. (2007). Reconsidering the role of structure in vision. In A. Markman & B. Ross (Eds.). *Categories in use series: The psychology of learning and motivation* (47, pp. 157–180). San Diego, CA: Academic Press.

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

DeWinter, J., & Wagemans, J. (2006). Segmentation of object outlines into parts: A large-scale, integrative study. *Cognition, 99*, 275–325.

Edelman, S. (1993). Representing three-dimensional objects by sets of activities of receptive fields. *Biological Cybernetics, 70*, 37–45.

Hummel, J. E., & Stankiewicz, B. J. (1997). Categorical relations in shape perception. *Spatial Vision, 10*, 201–236.

Hoffman, D. D., & Richards, W. A. (1984). Parts of recognition. *Cognition, 18*, 65–96.

Huttenlocher, D. P., & Ullman, S. (1990). Recognizing solid objects by alignment with an image. *International Journal of Computer Vision, 5*, 195–212.

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

Keane, S., Hayward, W. G., & Burke, D. (2003). Detection of three types of changes to novel objects. *Visual Cognition, 10*, 101–127.

Le Grand, R., Mondloch, C. J., Maurer, D., & Brent, H. P. (2001). Neuroperception: Early visual experience and face processing. *Nature, 410*, 890.

Marr, D., & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London, B: Biological Sciences, 200*, 269–294.

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience, 2*, 1019–1025.

Riesenhuber, M., Jarudi, I., Gilad, S., & Sinha, P. (2004). Face processing in humans is compatible with a simple shape-based model of vision. *Proceedings of the Royal Society of London, B: Biological Sciences, 271*, 448–450.

Saiki, J., & Hummel, J. E. (1998). Connectedness and part-relation integration in shape category learning. *Memory and Cognition, 26*, 1138–1156.

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology, 46A*, 225–245.

Tarr, M. J., & Bülthoff, H. H. (1995). Is human object recognition better described by geon structural descriptions or by multiple views—Comment on Biederman and Gerhardstein (1993). *Journal of Experimental Psychology: Human Perception and Performance, 21*, 71–86.

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

Vanrie, J., Willems, B., & Wagemans, J. (2001). Multiple routes to object matching from different viewpoints: Mental rotation versus invariant features. *Perception, 30*, 1047–1056.

Wagemans, J., van Gool, L., Lamote, C., & Foster, D. H. (2000). Minimal information to determine affine shape equivalence. *Journal of Experimental Psychology: Human Perception and Performance, 26*, 443–468.

Wallis, G., & Rolls, E. T. (1997). Invariant face and object recognition in the visual system. *Progress in Neurobiology, 51*, 167–194.

Yovel, G., & Duchaine, B. (2006). Specialized face perception mechanisms extract both part and spacing information: Evidence from developmental prosopagnosia. *Journal of Cognitive Neuroscience, 18*, 580–593.

Zhang, L., & Cottrell, G. W. (2005). Holistic processing develops because it is good. In B. G. Bara, L. Barsalou, & M. Bucciarelli (Eds.), *Proceedings of the 27th annual cognitive science conference* (pp. 2428–2433). Mahwah, NJ: Erlbaum.