A Layered View of Shape Perception

E. J. Green, Rutgers University
Penultimate Draft (forthcoming in *The British Journal for the Philosophy of Science*)

Abstract

This article develops a view of shape representation both in visual experience and in subpersonal visual processing. The view is that, in both cases, shape is represented in a *layered* manner: An object is represented as having multiple shape properties, and these properties have varying degrees of abstraction. I argue that this view is supported both by the facts about visual phenomenology and by a large collection of evidence in perceptual psychology. Such evidence is provided by studies of shape discriminability, apparent motion, multiple-object tracking, and structure-from-motion. Recent neuroscientific work has also corroborated this psychophysical evidence. Finally, I draw out implications of the layered view for processes of concept acquisition.

1 Introduction

The ability to conceptualize objects in the scene before our eyes depends in large part on seeing their shapes. It is by seeing the shapes of cars, buses, and motorcycles that you are able to cognize them as cars, buses, and motorcycles, respectively. As such, the question of how visual perception presents shape properties to thought deserves close philosophical scrutiny. In this article I'll propose a view of how shape properties are represented both in visual experience and in subpersonal visual processing. My thesis is that, in both cases, shape is represented in a *layered* manner: An object is represented as having multiple shape properties, and these properties have varying degrees of abstraction. Call this the *layered view* of shape perception.

The plan for the article is as follows. In section 2, I introduce the distinction between a *metric* property and an *abstract shape* property. Roughly, metric properties depend essentially on certain distance and/or angular measurements, while abstract shape properties do not—they are more qualitative. In section 3, I discuss some views of shape perception in the psychological and philosophical literature. I suggest that on several psychological views, the visual system's *subpersonal representation* of shape is wholly metric (i.e., the visual system only explicitly encodes the metric properties of objects), and that on some philosophical views, the representation of shape in visual *experience* is wholly metric (i.e., only metric properties figure in visual shape phenomenology). In section 4, I argue that the visual experience of shape is layered, rather than metric. To preview, my argument is that such layering is necessary in order to explain patterns of salience in the differences among various shape experiences. In section 5, I discuss a host of evidence indicating that the visual system extracts and uses information about abstract shape in a variety of processing tasks. In section 6, I argue that such evidence vitiates the proposal that the subpersonal representation of shape is wholly metric, and weighs in favor of the view that the visual system encodes shape in a layered manner. In section 7, I discuss some evidence concerning the neural underpinnings of abstract shape perception. In section 8, I suggest that the layered view has important implications for the process of concept acquisition.

2 Metric Properties and Abstract Shape Properties

It is common to arrange shape properties according to their relative *stability*, where the stability of a shape property is given by its invariance under geometrical transformation (change).[1] On this construal, shape property *A* is less stable than shape property *B* iff the transformations under which *A* is invariant form a proper subset of the transformations under which *B* is invariant. Thus, for example, the property of being a square is less stable than the property of being a rectangle, which in turn is less stable than the property of being a quadrilateral.

If an object *o* has shape properties *A* and *B*, and *A* is less stable than *B*, then an asymmetric entailment holds between the two: *o*'s having *A* entails that it has *B*, but not vice versa. Thus, *o*'s being rectangular entails that *o* is quadrilateral, but the converse is not the case. Furthermore, if *o* has shape properties *A* and *B*, and *A* is less stable than *B*, then a transformation cannot cause *o* to lose *B* without also causing it to lose *A*. Thus, if *o* starts out as a rectangle and so as a quadrilateral, then any transformation that causes *o* to cease to be quadrilateral must also cause *o* to cease to be rectangular.

Formally, I'll define the notion of a *metric property* as follows: A property *F* is a metric property iff *F* is invariant only under some subset of the similarity transformations. The similarity transformations include translation (simple change of position), rotation, reflection (change in "handedness"), and uniform scaling (simple change in size). Less formally, we can think of metric properties as properties that fail to survive changes in distances, lengths, and/or angles. They include, for example, being a square (which depends on having four angles of exactly 90°), being a square with a 20-inch perimeter, and being a circle with a 10-foot radius. Metric properties also include features that are much less stable, such as an object's precise location within a frame of reference, which fails to survive even translation or rotation. Thus, one

[1] This general approach traces back to the mathematician Felix Klein's innovative work in the 1870s (known as the Erlangen program) on the stratification of geometries according to the relative stability of the properties they examine.

type of metric property that will be particularly important in what follows is the location of a visible surface patch within a frame of reference centered on the viewer (i.e., viewer-centered distance and direction). This property is invariant under *none* of the similarity transformations, so trivially it is invariant under a subset of them.

Correspondingly, I'll define the notion of an *abstract shape property* as follows: A property *F* is an abstract shape property iff *F* is invariant under some proper superset of the similarity transformations. As such, we can think of abstract shape properties as ones that *can* survive at least some changes in distances, lengths, and angles. Properties like being a parallelogram or being a triangle are abstract, because they survive stretching and shearing, both of which alter a figure's constituent edge lengths and angles.[2] For instance, suppose a parallelogram is stretched along its horizontal axis so that its top and bottom edges double in length. After this transformation, its edge lengths and angles are different, but it remains a parallelogram. Thus, abstract shape properties are more stable than metric shape properties, and asymmetric entailments obtain between the two—e.g., something's being square entails that it is a parallelogram, but not vice versa.

I'll concentrate on two types of abstract shape property here: *topological properties* and *affine shape properties*. A topological property is any property that is preserved under all topological (i.e., one-to-one, continuous) transformations. Topological transformations are often called "rubber sheet" transformations, because they include all the deformations one can apply to a rubber sheet—e.g., twisting, stretching, bending, etc. However, they do not include tearing an object in two, poking holes in an object, "filling in" the holes of an object, or "gluing" pieces of the object together. Topological properties include connectedness, an object's number of holes, and an object's property of being inside or outside another object. Because of the generality of topological transformations, any two solid figures—e.g., a ball and a block—are topologically equivalent.

An affine shape property is any property that survives affine transformations. Roughly, affine transformations include the similarity transformations along with stretching and shearing

[2] In a stretch transformation, all points are moved in a direction perpendicular to a fixed axis, and move by an amount proportional to their initial distance from the axis. In a shear transformation, all points of an object are moved in a direction parallel to a fixed axis, and move by an amount proportional to their initial distance from the axis. A shear transforms, e.g., a rectangle into a (non-rectangular) parallelogram.

along an arbitrary direction.[3] Affine shape properties include: collinearity, being straight vs. curved, parallelism, ellipticality, triangularity, being a parallelogram, coplanarity of lines, the number of sides in a polygon, and signs of curvature (concave vs. convex) along the surface of an object (Todd [2004]).[4] Since distances and angle magnitudes are not preserved under affine transformation, affine shape properties are more stable than metric properties. Thus, if one surface is a stretching of another surface, then the two are affine equivalent, even though they are metrically distinct. Moreover, since the affine transformations form a subset of the topological transformations, it follows that any topological property also counts as an affine shape property. However, when I refer here to affine shape properties, I'll have in mind properties that are affine invariant but *not* topologically invariant, such as those listed above. Figure 1 shows examples of topological transformation and affine transformation.

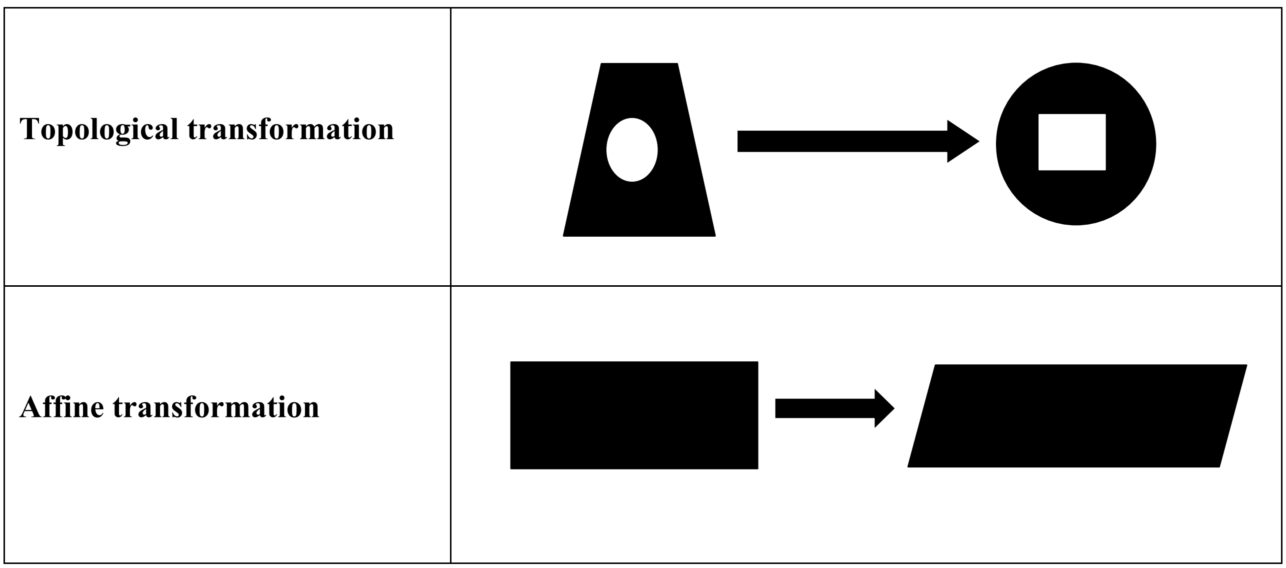

Figure 1. Examples of topological transformation and affine transformation

[3] Formally, an affine transformation is a function $f(x) = \mathbf{A}x + \mathbf{b}$, such that $\mathbf{A}$ is an invertible matrix, x is a coordinatized point, and $\mathbf{b}$ is a position vector. Affine transformations thus include linear transformations with the addition of translation.

[4] A helpful way to visualize the types of changes possible under affine transformation is via the close relation between affine transformation and parallel projection: Any affine transformation is equivalent to a composition of at most two parallel projections (Brannan *et al.* [2012], p. 84). Moreover, many such transformations (though not uniform scaling) can be expressed as a single parallel projection. Thus, affine transformations can be visualized by imagining a parallel projection mapping one plane to another. If a figure A is specified on the plane that is the preimage of the mapping, then the figure on the projection plane will differ from A by at most an affine transformation.

3 Metric Views

Many have been committed to what I'll call *metric views* of either visual *representation* (at the subpersonal processing level) or visual *experience*. This section introduces these positions, in preparation for arguing against them.

I'll construe a *metric representation* of shape as one that only explicitly encodes metric properties, such as locations, distances, lengths, and angles. I won't attempt to offer a reductive analysis of the notion of explicit representation here, but the notion can be intuitively cashed out as follows. When a representation makes certain information explicit, that information is made immediately available for use by the system that uses the representation. By contrast, when a representation leaves certain information implicit, further computations are necessary in order to extract that information (Kirsh [2003]). An illustration of this difference is due to David Marr ([1982], p. 20): The Arabic numerical system makes explicit a number's decomposition into powers of ten (e.g., "63" in the Arabic system is equal to $6*10^1 + 3*10^0$), while leaving its composition into powers of two implicit. The binary numerical system, on the other hand, makes explicit a number's decomposition into powers of two (e.g., "1011" in the binary system is equal to $1*2^3 + 0*2^2 + 1*2^1 + 1*2^0$), while leaving its decomposition into powers of ten implicit.

A representative kind of metric representation is Marr's 2½-D sketch, which can be construed as a type of depth map. It is an array specifying the viewer-centered distance, direction, and local orientation at each point (up to a certain resolution) for all visible surfaces in the scene (see Marr [1982], pp. 275-9).[5] The 2½-D sketch is a metric representation because it only explicitly encodes viewer-centered locations and angles (specifically, the locations of small surface patches and the angles of surface normals relative to the line of sight[6]), and these are metric properties.[7] Moreover, substantial computation is needed in order to extract (most) non-metric properties on the basis of a 2½-D sketch. (This is also partly because of the *local* character of the 2½-D sketch—see note 7.) For instance, to extract the abstract shape property

[5] Because the 2½-D sketch is limited to describing the geometry of *visible* surfaces, it does not include any description of the way surfaces complete behind occluders.

[6] A surface normal at point *p* on a surface is a line segment perpendicular to the plane that is locally tangent to (i.e., "just grazes") the surface at *p*. Marr proposed that local surface orientation at a point *p* is specified in the 2½-D sketch by encoding the angle formed by the surface normal at *p* and the viewer's line of sight.

[7] The 2½-D sketch representation is also *local*—geometrical properties (e.g., location, orientation) are ascribed only to very small elements of the scene, such as small surface patches and edge segments. Some subsequent theorists have rejected the local assumption (see Jackendoff [1987], pp. 331-8), suggesting extensions of the 2½-D sketch that explicitly segment the scene into objects, surfaces, backgrounds, etc.

‘parallelogram’ on the basis of a 2½-D representation of a surface, the system must perform computations to verify, *inter alia*, that the surface has *four straight edges*, that those edges are *connected*, and that two pairs of those edges are *parallel*. None of this information is made explicit by the 2½-D sketch—indeed, the 2½-D sketch doesn’t even have the resources for representing parallelism or number of sides. Thus, if the 2½-D sketch encodes abstract shape properties at all, it does so only implicitly.

Call a view on which the visual system represents shape only via metric representation a *metric view* of visual shape representation. Marr himself did not hold a metric view. In Marr’s theory, the 2½-D sketch was followed by a 3-D structural description, which represents certain abstract shape properties of objects and their parts (see Marr [1982], ch. 5; Marr and Nishihara [1978]; see also Biederman [1987], [2013]).[8] Thus, the 3-D model might simply represent an object’s part—say, a person’s leg—as “roughly cylindrical.” Nonetheless, while Marr clearly thought that visual shape analysis was not exhausted by the 2½-D sketch, several subsequent accounts of visual shape representation—primarily in the object recognition literature—have closely resembled the 2½-D sketch in important ways.[9] The most common position in this vein is the so-called *view-* or *image-based* approach (Tarr and Pinker [1989]; Ullman and Basri [1991]; Ullman [1996], [1998]; Edelman [1997], [1999]; Williams and Tarr [1999]; Riesenhuber and Poggio [2002]; Graf [2006]). According to most such proposals, the visual system represents an object’s shape simply by specifying the numerical coordinates of certain local features of the object (or the object’s projected image).[10] For instance, on Ullman’s ([1996]; [1998]) approach,

[8] It is interesting to observe, however, that Marr called the 2½-D sketch “the end, perhaps, of *pure* perception” ([1982], p. 268, emphasis added). Steven Pinker ([1997], p. 260) appears to endorse a similar view. It is unclear what Marr had in mind by “pure perception.”

[9] Vision scientists *outside* the object recognition literature have often rejected the metric view of shape representation. Notable opponents to the metric view include of course the researchers whose work is discussed below (e.g., Biederman, Chen, Koenderink, Todd, and Wagemans). Furthermore, many psychologists working on perceptual organization have placed emphasis on the perceptual recovery of affine properties such as collinearity and parallelism, and topological properties such as closure and connectedness, since these are important cues to perceptual grouping, figure-ground segregation, and/or amodal completion (see, for example, Feldman [2007]; Hoffman [1998]; Kellman [2003]; Nakayama and Shimojo [1992]; Palmer [2003]; and Tse [1999]; see Wagemans *et al.* [2012] for review). Vision scientists who study the processes of extracting shape from line drawings, shading, or texture also generally reject the metric view, sometimes in favor of a view on which vision represents affine shape (see Belhumeur *et al.* [1999]; Cole *et al.* 2009; Koenderink *et al.* [2001]); Phillips *et al.* [2003]; Todd [2004]).

[10] Though all such views agree that the coordinate system in which features are specified is viewer-centered, they differ on whether it is 2-D (Ullman [1998]; Edelman [1999]) or 3-D (Williams and Tarr [1999]). Ullman ([1996], pp. 110-2) suggests that depth values are used when they are available, but the model he adopts does not require them. This distinction between 2-D and 3-D view-based schemes will not matter for current purposes, since either type of representation is metric in nature.

the representation of shape that serves as input to object recognition is a vector specifying the viewer-centered 2-D locations of simple image elements like edges, vertices, and contour inflection points. An input vector *v* of this sort is recognized as deriving from a particular object *o* if the visual system can obtain *v* by linear combination of a small number of vectors stored in memory that are known to correspond to distinct images of *o*.[11] This proposal, and others like it, shares a critical feature with the 2½-D sketch—namely, it entails that vision only explicitly represents certain metric features of objects, such as the viewer-centered locations of their edges, vertices, etc. (after normalizing for position, rotation, and scale). Transformations of an object outside of the similarity group (e.g., stretching, shearing, or bending) will alter these locations.

The metric view of visual shape representation is a theory about subpersonal visual processing. However, it suggests a counterpart in the domain of phenomenology, which I'll call the metric view of visual shape *experience*. On this proposal, the only geometrical properties represented in visual experience are metric properties, such as the locations and orientations of small surface patches.

Although detailed theories of shape phenomenology are relatively scant in the philosophical literature, the metric view can be found in some authors. For instance, certain passages suggest that Evans ([1985]) endorsed a version of this view. Evans proposes that visual experience represents shape solely by egocentrically locating the points of visible surfaces in "behavioral space"—a specification of space common to each modality. Thus: "To have the visual experience of four points of light arranged in a square amounts to no more than being in a complex informational state which embodies information about the egocentric location of those lights" (Evans [1985], p. 339).[12] This indicates that, for Evans, the experience of shape amounts, roughly, to representing the viewer-centered locations of visible surface points.[13]

Peacocke's ([1992]) notion of *scenario content* bears some similarity to Evans's proposal. According to Peacocke, at the most fundamental level, visual experience represents a

[11] Ullman and Basri ([1991]) proved that, under certain conditions (e.g., when an object is rigid, all its points are visible in each view, and points are correctly "matched" across images), the vectors of X- and Y- coordinates of points in a specific image of an object (under parallel projection) can be expressed as linear combinations of such vectors in three distinct images of the same object.
[12] Page references for Evans ([1985]) correspond to the reprint found in Noë and Thompson ([2002]).
[13] A caveat: Evans does not explicitly claim that the approach to shape experience that he endorses for configurations of points of light also holds for experiences of solid figures or surfaces. Thus, it is possible that he would have rejected the metric view as a complete account of shape perception.

positioned scenario. This is described as a way of filling out space relative to an origin and axes fixed on the center of gravity of the perceiver's body ([1992], p. 63). More specifically:

> In picking out one of these ways of filling out the space, we need to do at least the following. For each point…identified by its distance and direction from the origin, we need to specify whether there is a surface there and, if so, what texture, hue, saturation, and brightness it has at that point, together with its degree of solidity. The orientation of the surface must be included. So must much more in the visual case: the direction, intensity, and character of light sources; the rate of change of perceptible properties, including location; indeed, it should include second differentials with respect to time where these prove to be perceptible. ([1992], p. 63)

Again, this essentially amounts to a point-by-point representation of surface depth and orientation (though other local features are included as well). And as such, scenario content specifies, in the first instance, metric properties—point-wise distance, direction, and orientation relative to the viewer. However, Peacocke recognizes the need to enrich the scenario content approach in order to account for certain well-known perceptual phenomena.[14]

The layered view is consistent with (but does not entail) the view that visual experiences have scenario content. But if the layered view is right, visual experiences must also have much *more* than scenario content. In particular, visual experiences must represent abstract shape properties in addition to metric properties.

Finally, I should note that the metric view of visual shape experience is quite pretheoretically attractive. It is natural to think of visual experience as simply delivering a pixilated map of the environment specifying the distances and directions of individual surface points. On this picture, it is the job of cognition to "carve up" this map in certain ways and extract abstract shape categories (e.g., triangle, quadrilateral, solid figure, etc.) on that basis.[15]

---

[14] Consider, for instance, Mach's tilted-square/regular-diamond figure, which can be seen either as a diamond or as a tilted square. Both of these percepts are compatible with precisely the same scenario content (e.g., viewer-centered distance, direction, and orientation). Because of this, Peacocke introduces a further layer, which he calls "protopropositional content." Protopropositional content is truth-evaluable, and it consists of individuals, properties, and relations. Peacocke suggests that protopropositional content includes the properties of being square, diamond, collinear, curved, parallel, and symmetric. The layered view is consistent with this proposal, though it is consistent with other views as well.

[15] Though this is not the place for historical exegesis, it is worth asking whether Berkeley ([1710/1982]) held a metric view of visual shape experience, given his famous rejection of the "abstract general idea" of triangularity ([Introduction, §13]). Though Berkeley allows that a determinate idea of a particular triangle may on occasion function to "stand for and represent" the property of triangularity ([Introduction, §15]), he seems to have believed that this requires a cognitive act on the part of the subject—one must *use* the idea in a certain way. Thus, though the matter is by no means clear-cut, it seems fair to assume that Berkeley would have agreed that visual experience itself presents us only with determinate metric properties.

In what follows I will argue against both types of metric views. I'll first argue that the layered view of visual shape experience does a better job than the metric view of explaining patterns of salience in the differences among shape experiences. Then I'll argue that the metric view of visual shape representation cannot explain the visual system's ability to put information about abstract shape properties to use in a number of processing tasks.

4 Against Metric Views of Visual Shape Experience

This section addresses the question of which geometrical properties are represented in the *conscious visual experience* of shape. For present purposes, I will simply assume that visual experiences attribute properties to objects in the environment, and that the representation of such properties makes a difference to visual phenomenology.

As noted above, the metric view of shape experience holds that visual experience presents us *only* with metric properties of objects, such as point-wise distance and orientation. Another view—the one I'll defend here—is that states of visual experience represent geometrical properties at multiple levels of abstraction simultaneously. Call this the *layered view* of visual shape experience. Thus, for example, when viewing a triangular surface, you might simultaneously experience it as: (i) a surface composed of points located in such-and-such a direction, at such-and-such a distance, and at such-and-such an orientation relative to your line of sight, (ii) a triangle, and (iii) a solid figure.

How can we determine which (if either) of these views is correct? Perhaps the most obvious way would be to simply introspect one's experience and see whether it reveals the representation of abstract shape properties in addition to metric properties. Unfortunately, however, the method of introspection faces a number of well-known problems (see Schwitzgebel [2011]). Moreover, if I introspect my experience and claim to encounter abstract shape properties while you introspect yours and claim to encounter only metric properties, how can we determine who is right?

A more promising option, it seems, would be to employ the method of *phenomenal contrast*, recently championed by Susanna Siegel ([2010]). This method works as follows. First, formulate the hypothesis that a property *F* is represented in visual experience. Next, examine two overall experiences, *A* and *B*, such that (i) *A* is a candidate for representing *F*, (ii) *B* is not such a candidate, and (iii) *A* and *B* are as similar as possible in other respects. Then check whether *A* and *B* contrast phenomenally. If they do, then determine whether the proposal that *A* includes a visual experience that represents *F* provides the best explanation of this phenomenal contrast. Critically, this last stage can invoke empirical considerations (see Block [2014]), though Siegel does not generally do so.

Siegel uses the phenomenal contrast strategy to defend the view that certain "high-level" properties, such as causation, natural kinds, etc., are represented in visual experience, alongside the usual suspects (color, shape, motion, etc.). Thus, for evaluating this hypothesis, the method of phenomenal contrast recommends that we examine two experiences that are essentially identical in respect of the colors, shapes, etc., that they represent, but perhaps differ in the representation of such high-level properties. If the two experiences differ phenomenally, then (perhaps!) the best explanation is that one represents high-level properties while the other does not.

In the current case, we wish to compare the hypothesis that visual experiences *only* represent metric properties with the hypothesis that visual experiences *also* represent abstract shape properties. Thus, the most straightforward application of the method of phenomenal contrast would be to examine two experiences that are essentially identical in the metric properties that they represent, but perhaps differ in their representation of abstract shape properties. Any difference in shape phenomenology between these two experiences could be taken to support the layered view.

But now we face a problem. Given that any difference in abstract shape entails *some* difference in metric properties (as discussed in section 2), it is *prima facie* plausible that any change in the visual experiential *representation* of abstract shape will entail *some* change in the visual experiential *representation* of metric properties.[16] Accordingly, any pair of phenomenally contrasting experiences that even potentially differ with respect to their representation of a particular abstract shape property (e.g., triangularity) will also plausibly differ with respect to their representation of numerous metric properties (e.g., locations of surface points, edge lengths, angles, etc.). So it seems unlikely that we will be able to find two experiences that uncontroversially agree in their representation of metric properties, but perhaps differ in their

[16] I do not actually endorse the latter entailment, and in fact I suspect that it does not hold (though I admit that it has pretheoretic plausibility). But I aim to show here that even if we grant the entailment, we still have strong reasons to suppose that abstract shape properties are represented in visual experience.

representation of abstract shape properties. So how can we identify a pair of experiences that allow us to appropriately compare the two hypotheses of interest?[17]

This is a tricky situation, but I suggest that there is a maneuver available. Rather than looking merely at two individual shape experiences, we can examine *pairs of changes* in visual shape phenomenology, one of which clearly involves a change *only* in the representation of metric properties, and the other of which is a candidate for *also* involving a change in the representation of a given abstract shape property. The hypothesis recommended by the layered view is that, other things being equal, changes of the latter type should be *more salient* (i.e., more noticeable or striking) than changes of the former type.

However, on any view of shape experience—including the metric view—certain shape changes should be expected to be more salient than others. For instance, a transformation that stretches a rectangle by a factor of 2 should be more salient than a transformation that stretches it by a factor of 1.5, simply because, e.g., point locations are altered more in the former case. Thus, the claim is *not* that the metric view cannot predict that certain changes will be more salient than others—trivially, it can. Rather, the claim is that the layered view offers a *better* explanation of the *specific* patterns of salience associated with shape changes. This is because the metric view on its own does not predict that the salience of a given shape change should be sensitive to whether or not that change crosses the boundary of an abstract shape category. The layered view, on the other hand, does predict this.

One method, then, would be the following: First formulate a hypothesis about the experience of abstract shape—e.g., "Some visual experiences represent abstract shape property *F*." Then consider the visual experience of a *base stimulus* that has *F*. Next, consider experiences of two stimuli—which we can call *test stimuli*—that meet the following conditions: Both test stimuli differ from the base stimulus in their metric properties, but test stimulus 1 shares *F* with the base stimulus, while test stimulus 2 does not. According to the layered view of visual shape experience, the experience of the base stimulus differs from the experience of test stimulus 2 as regards the representation of (at least) *two types* of properties—both metric properties and

[17] This problem in applying the method of phenomenal contrast is liable to arise in any situation where one wishes to compare two hypotheses, *P* and *Q*, where *P* claims that only *determinates* within a particular category (e.g., scarlet) figure in conscious experience, while *Q* claims that *determinables* within that category (e.g., red) *also* figure in conscious experience. Moreover, I suspect that analogs of the method discussed next—viz., examining patterns of salience among changes in experience, rather than simply comparing two individual experiences—for overcoming this difficulty may be applied in other cases.

abstract shape property *F*—while it differs from the experience of test stimulus 1 only in the representation of metric properties. As such, the layered view would predict that—other things being equal—the former difference will be more salient than the latter.

Holding other things equal, however, is no easy task. Roughly, we want to ensure that the two changes (base stimulus to test stimulus 1 vs. base stimulus to test stimulus 2) are approximately comparable, aside of course from the critical geometrical difference (viz., one crosses the boundary of a relevant abstract shape category, while the other does not). Most importantly, we want to ensure that any difference in the salience of the two changes has to do with perception of abstract shape properties, rather than with detecting differences in very local features, such as point or pixel locations. Perhaps more intuitively, we want to ensure that the change from the base stimulus to test stimulus 2 isn't more salient simply because the two figures have less "overlap" in their constituent points than the base stimulus and test stimulus 1.

Psychologists and computer scientists who have faced this problem have developed measures of the degree to which two stimuli overlap in their local features (see Veltkamp and Latecki [2006]). Thus, suppose that we represent figures within a coordinatized frame of reference. A given figure can then be represented by a binary vector indicating, for each point *p* within the reference frame, whether *p* "belongs" to the figure ("1" if the point belongs, "0" if it does not). Given this scheme, one simple way to measure the difference between the overall "point distribution" of two figures (and thus the change between them), called the Hamming distance (Ullman [1996], p. 5), would be to first normalize two figures to a standard position and orientation, then add up the number of places in which the vectors for the two shapes differ. Another method would be to find, for each point *p* belonging to one figure, the distance from *p* to the closest point belonging to the other shape, and take the maximum of these distances (known as the Hausdorff metric). Yet another method would be to simply sum the distances between each point of one figure and its nearest neighbor in the other figure, which would give a measure of the overall "point displacement" from one figure to the other (again, following normalization).[18]

While it is fortunate that such measures exist, their disparateness may make it seem impossible to "hold other things equal" across two shape changes. Nonetheless, there is a way

[18] Alternatively, we might sum the *squared* distances between corresponding points of the two figures, and take the square root of this sum (known as the Procrustean distance).

forward. We can ensure that, no matter which of these measures is used, test stimulus 2 is at least as—if not more—different from the base stimulus in its local features. And luckily, stimuli that obey this restriction have been used in a recent shape discrimination study by Todd, Weismantel, and Kallie ([2014]). Todd *et al.* set out to compare the detectability of shape changes at varying levels of abstractness. Subjects were first shown a stimulus—the base stimulus—for 300 ms. Then, after a brief delay, they were shown two other stimuli in succession, each for 300 ms. One of these stimuli was metrically equivalent to the base stimulus, while the other was metrically distinct. The subjects' task was simply to indicate which of these two objects was equivalent to the base stimulus. The metrically distinct stimulus could differ from the base in one of four ways: It could involve a stretching (change in contour length), a skewing causing very slight convergence of contours that were parallel in the base stimulus (loss of parallelism), the addition of a bump to the base stimulus's contour (loss of collinearity), or the introduction of a hole (change in topology). The first type of change disrupts only metric shape, leaving affine shape and topology intact. The second and third types of changes disrupt affine shape, but leave topology intact.[19] The fourth type of change disrupts topology. Examples of these changes are shown in figure 2.

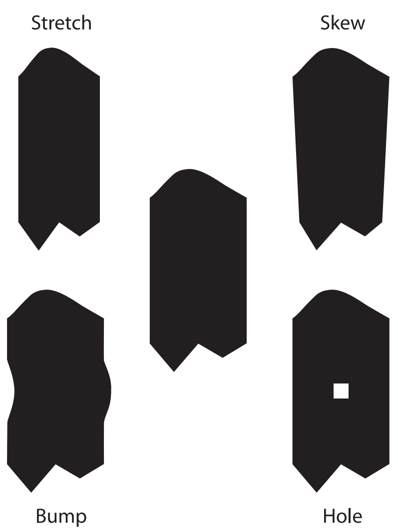

Figure 2. Examples of the changes used in Todd et al. (2014). The base stimulus is shown in the center. Reproduced with permission of the Association for Research in Vision and Ophthalmology.

[19] Loss of collinearity also alters an object's *projective* properties. A projective property is a property that is preserved under projective transformations. Since the affine transformations form a subset of the projective transformations (Brannan *et al.* [2012]), any property that is preserved under all projective transformations is also preserved under all affine transformations. Thus, every projective property is an affine property, but not vice versa. For present purposes, I focus on the larger set of affine properties, but it is possible that the two types of properties have different degrees of salience in visual phenomenology. Indeed, I find loss of collinearity to be more phenomenologically salient than loss of parallelism. This is borne out in the results of Todd *et al.* ([2014]).

Before covering the results of this study, I recommend that you consider your experiences of the stimuli in figure 2, and try to decide which changes are most phenomenologically salient. For me at least, the result is fairly clear. The changes that disrupt abstract shape (skewing, adding a bump, or adding a hole) are more salient than the change that disrupts only metric shape (stretching). Indeed, they strike me as 'qualitative' in a way that the latter change does not, even though the overall point displacement (for instance) is actually greater in the stretching transformation. This should make initially plausible the view that abstract shape properties (e.g., parallelism, collinearity, and number of holes) figure in shape phenomenology alongside metric shape properties (lengths, angles, and curvature).

The results of the experiment comported with this intuition.[20] Todd *et al.* analyzed subjects' performance in cases where—by almost all common measures of local feature differences between figures, such as those discussed above—the topologically distinct stimulus was *less different* from the base than the affine distinct stimuli, and the affine distinct stimuli were in turn less different from the base than the merely metrically distinct stimulus. It was found that, even in these conditions, subjects were better at performing the task in the affine change conditions (when one of the stimuli involved skewing or adding a bump to the base stimulus) than in the mere metric change condition (stretching), and were better still in the topological change condition (addition of a hole). This indicates that, given two shape changes *A* and *B* such that (i) *A* disrupts an abstract shape category while *B* does not, and (ii) by all or most available measures, the magnitude of local feature difference is either roughly comparable or somewhat greater in the case of *B*, *A* tends to be more salient than *B*.[21]

Now, assuming that subjects perform discrimination tasks like this on the basis of their *visual* shape phenomenology (whether this is the case will be considered shortly), these results raise a challenge for metric views of visual shape experience. For if visual experience does *not*

[20] Moreover, this is by no means the only study to document increased salience for changes that cross the boundary of an abstract shape category. See also the study by Kayaert and Wagemans ([2010]) described below, along with Amir *et al.* ([2014]), Biederman and Bar ([1999]), and Todd *et al.* ([1998]). Comparisons across these studies are admittedly difficult, however, because slightly different measurements of local feature differences across figures were used.

[21] Condition (ii) is crucial, of course. If the metric change (stretching) were made very extreme (e.g., compressing the object to only a few pixels) then it would likely be more salient than the changes in affine shape or topology shown above. But this is not a problem for the layered view. On the layered view, the salience of a particular shape change is predicted to be a complex function of differences in geometrical properties at varying degrees of abstraction—*including* metric properties. As such, if the change in metric properties (lengths, angles, point locations, etc.) is extreme enough, then it should be expected to be more salient than a given transformation that disrupts abstract shape, if the change in metric properties in the latter case is much smaller.

represent abstract shape properties, and instead only represents, e.g., the viewer-centered locations of surface points, then we have no obvious explanation of why changes between objects that alter abstract shape should be especially salient in visual shape phenomenology. But the layered view offers a natural explanation for this.

I'll now consider two potential responses on behalf of the metric view.

First, one might suggest that a version of the metric view could predict the results of the Todd *et al.* experiment without invoking the representation of abstract shape properties if the view simply posited an appropriate subjective similarity function *R* over experiences of metric properties (assuming that discriminability tracks subjective similarity). That is, perhaps experiences represent *only* metric properties such as length, distance, location, and angle, but, by *R*, experiences of metrically distinct but affine equivalent objects turn out to be (other things being equal) more subjectively similar than experiences of affine distinct objects. (Of course, however, *R* could not be based on any of the measures of local feature difference given above.)

I suspect that an appropriate subjective similarity function could indeed predict the results of the Todd *et al.* study (though to have general applicability the measure would likely need to be forbiddingly complex and context-sensitive[22]). Note, however, that a similarity function *R* over shape experiences would be compatible with *either* the metric view *or* the layered view of the contents of shape experiences. But upon reflection, I think we still have reason to favor the layered view, because the layered view offers a better account of why *R* is the "right" indicator of similarity between shape experiences. On the layered view, the reason why—other things being equal—experiences of objects within the same abstract shape category are subjectively more similar than experiences of objects from different abstract shape categories is because the latter objects are represented in experience to *differ* in that shape category, while the former are not represented to so differ. The metric view, on the other hand, does not have any ready explanation of what *grounds* these facts about subjective similarity. Rather, on the metric view the relevant similarity function would be left brute and unexplained.

A second response for the metric view is recommended by closer attention to the view-based models of object recognition discussed above. According to several view-based models, the representation of shape is *sparse*—it involves simply representing an *n*-tuple composed of

---

[22] There are, it should be noted, numerous factors that seem to affect how similar two shapes are seen to be. One important contributor, which I will not discuss here, is whether two shapes can be decomposed into parts with roughly the same metric structure (e.g., Barenholtz and Tarr [2008]).

the viewer-centered coordinates of "critical features" like vertices, edges, curvature extrema, and inflection points. Perhaps, then, visual experience is sparse in the same way—only the coordinates of such critical geometrical features are represented. Now, importantly, some of the changes in abstract shape used by Todd *et al.* ([2014]) (viz., adding a bump or a hole), involved adding *extra* vertices or curvature extrema. As such, on some view-based proposals, this would result in the addition of extra elements to the *n*-tuple specifying object shape. Stimulus stretching, on the other hand, did *not* involve adding an extra critical feature. Perhaps, then, it will be claimed that—other things being equal—changes that result in the addition or subtraction of critical features are more experientially salient than changes that do not. This would explain why some of the abstract shape changes were more salient than the stretching change.

There are a couple of things to note in response. First, observe that the skewing change did *not* increase the number of vertices or curvature extrema in the object, yet was still more salient than stretching. Second, other studies have shown independently that models on which object shape is encoded simply as an *n*-tuple of critical feature coordinates do a relatively poor job of explaining patterns of salience in shape discrimination. Generally, such views predict that the dissimilarity of two shapes should be a function of the distances between their critical feature coordinates. However, studies specifically testing this prediction have found that discriminability is instead more strongly influenced by abstract shape properties of objects (e.g., whether the objects' axes are straight vs. curved) and abstract (or "categorical") relations among parts of the overall shape, such as whether one part intersects another part above vs. below the latter part's midpoint (e.g., Hummel and Stankiewicz [1996]; Biederman and Bar [1999]). Still, the issues in this area are complicated, so I leave open whether a "sparse" metric view may be able to predict the specific patterns of discriminability found in Todd *et al.* ([2014]).[23]

I noted above that studies of shape discriminability seem to favor the layered view, but only on the assumption that subjects perform such tasks on the basis of *visual phenomenology*. However, this assumption may be questioned. Perhaps the difference in salience is rooted not in visual experience, but rather in the way shape properties are cognized. Doubtless we generally

[23] Nevertheless, the evidence discussed in the next section and the arguments in section 6 provide, I think, strong reasons to doubt that the view-based approach (including the sparse versions of this approach) can provide a complete account of shape representation at the subpersonal level, although it may give part of the story. If these arguments succeed, then the defender of the metric approach to shape *experience* would then be in the position of explaining why *only* the metric components of subpersonal shape representation subserve visual phenomenology, while other components do not.

categorize objects in thought according to abstract shape properties (e.g., their number of sides), so perhaps shape changes are especially salient when they are accompanied by differences in postperceptual categorization.

Though it is quite difficult to conclusively rule out an alternative explanation of this sort, there are reasons to be skeptical of it.

First, the difference in salience between changes that preserve certain abstract shape properties (e.g., parallelism or solidity) and those that do not simply *feels* perceptual, rather than cognitive. Plausibly, the stimulus with a bump *visually appears* more different from the base stimulus than does the stretched stimulus. This does not seem to be a matter merely of how those stimuli are grasped in cognition.

Moreover, roughly the same patterns of salience have also been obtained with young infants. Kayaert and Wagemans ([2010]) used a dishabituation paradigm to study affine shape perception in infants and toddlers. The children were repeatedly shown either a triangle or trapezoid. After they habituated to this stimulus, they were presented with a display containing two test stimuli: (i) an object that differed from the original by only a metric change, and (ii) one that differed in affine structure (see figure 3). The former was a stretching of the habituation stimulus (but preserved its number of sides), while the latter transformed it either from a triangle into a trapezoid or vice versa. However, these two changes were constrained such that either they involved the same overall point displacement, or the change that preserved affine shape involved a larger difference than the change that failed to preserve affine shape. It was found that even the youngest infants (approximately 3 months) looked significantly longer toward the affine-distinct stimulus than the merely metrically-distinct stimulus, and the size of this effect did not differ significantly between younger and older children.

Figure 3. Stimuli used by Kayaert and Wagemans ([2010]). The triangle on the left differs from the triangle in the middle by a mere metric change (stretching) that preserves affine shape, while it differs from the trapezoid on the right in its affine shape. *Source*: Kayaert & Wagemans ([2010]).

The fact that abstract shape changes are already more salient (other things being equal) in infancy lends some support to the view that this contrast in salience is rooted in perception, because it suggests that the tendency to experience changes in abstract shape as more salient is present very early and is likely involuntary. These are hallmark features of a perceptual process (Fodor [1983]; Pylyshyn [1999]). Nevertheless, it should be admitted that this evidence is confirmatory, but not conclusive.

However, there is also a large amount of evidence that information about abstract shape is extracted and put to use in a number of *paradigmatically visual* processes, such as apparent motion perception, structure-from-motion, and object tracking. I contend that this, in conjunction with the above observations, provides good reason to believe that abstract shape properties are represented in *visual* experience, and not just in postperceptual phenomenology. I discuss this evidence next.

5 The Visual System Uses Abstract Shape Properties

There is now a great deal of evidence that both topological and affine properties play an important role in visual processing.[24] I begin with topological properties.

One way to test whether a property is extracted during early visual processing, rather than in, say, postperceptual cognition, is to use very short presentation times (e.g., Sekuler and Palmer [1992]). The idea is that early removal of a stimulus "interrupts" the processing of that stimulus. Thus, to probe for the perception of topological properties, Lin Chen ([1982]; [1990]) gave subjects a discrimination task in which they were shown pairs of figures for just 5 milliseconds and then asked to indicate whether the figures were the same or different in shape. In one experiment, the pair of figures was drawn from the following set: solid square, solid circle, ring, or solid triangle (see figure 4). Crucially, while the circle and the ring are very similar with respect to local metric properties such as contour curvature, they have different topologies (viz., one figure has a hole while the other does not). On the other hand, the solid circle is topologically equivalent with both the solid square and the solid triangle. The crucial measure was the rate of correctly reporting that two figures of different types were in fact different in shape. For if the visual system represents the topological properties of objects (such

[24] For an overview of evidence in favor of the perception of topological properties, see Chen ([2005]). For overviews of evidence in favor of the perception of affine properties, see Todd ([2004]) and Bennett ([2012]).

as their number of holes), then these properties should serve to distinguish the ring from the other figures, but should not serve to distinguish the other figures from one another. As such, one might expect to find *better* discrimination performance in the case of, say, the ring and solid circle than in the case of, say, the solid circle and triangle. And this was indeed found: subjects were significantly better at distinguishing the ring and solid circle (64.5% correct) than either the solid circle and square (43.5% correct) or the solid circle and triangle (38.5% correct). Moreover, this pattern of results continued to hold after differences in spatial frequency, luminous flux (i.e., the total amount of light energy provided by the figures), and area were held constant across topologically equivalent and topologically distinct pairs of stimuli (see Chen [1990]).

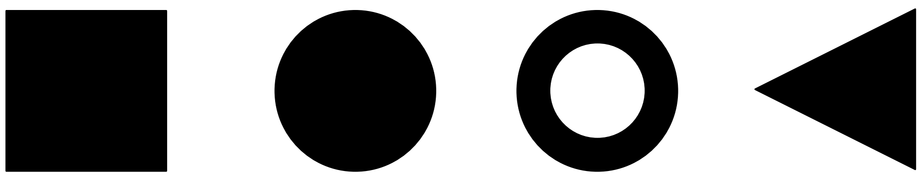

Figure 4. Stimuli similar to those used by Chen ([1982])

Chen has also examined the role that topological properties play in the perception of apparent motion. As is well known, when one stimulus is flashed and then another is flashed in a different location, then with a suitable spatiotemporal gap between the flashes, the viewer will have a visual experience as of a single object moving continuously from one location to the other. One interesting variant on this paradigm involves presenting multiple stimuli, rather than one, in the second frame. For example, the first stimulus *A* may be followed by a pair of stimuli *B* and *C*. In this case, the visual system faces the problem of "choosing" whether to represent motion from *A* to *B*, from *A* to *C*, from *A* to both *B* and *C* (i.e., "splitting"), or no motion at all. A heavily examined issue concerns which properties the visual system exploits in solving this problem (see Green [1986]). If a property is exploited in determining matches in apparent motion, this provides good evidence that the property is represented in vision, because it indicates that the visual system uses representations of the property during motion processing.[25] Thus, Chen ([1985]) showed subjects two frames in succession. In frame 1, a single stimulus occupied the center of the display, and in frame 2, the stimulus was replaced by two stimuli—one to the left of center and the other the same distance to the right of center. Subjects were

[25] It should be noted, however, that visual motion perception is not a single process, but rather involves a number of different subsystems (see Lu and Sperling [2001]).

asked to choose whether they saw motion from center to left or from center to right. For a wide variety of stimuli (and, again, with other differences between the stimuli controlled), subjects were significantly more likely to see motion from the central stimulus to a topologically equivalent stimulus than to a topologically distinct one. Thus, for example, if frame 1 contained a square with a square-shaped hole and frame 2 contained both a solid square and a ring, subjects were significantly more likely to see motion to the ring than to the solid square.

The proposal that topology is used in determining object identity over time has also been verified by a recent multiple object tracking study (Zhou *et al.* [2010]). Subjects were asked to keep track of four stimuli as they moved about the screen in the presence of a set of distractors. The stimuli could undergo various sorts of feature changes during a trial. The critical measure was how such changes impacted subjects' ability to keep track of the stimuli. It was found that changes in topology—though not other feature changes (e.g., changes in color or metric shape)—significantly impaired the ability to track an object over time, indicating that the visual system relies heavily on topology in order to determine whether an object has remained the *same* object from one moment to the next.

Each of these studies provides evidence that the visual system treats objects (or time slices of an object) that are very different in metric properties as nevertheless having certain features in common—namely, topological features. And moreover, the visual system apparently *uses* this information in certain processing tasks, such as, e.g., motion computations.

I turn now to the perception of affine shape properties.

Much of the work on affine shape perception has been motivated by a large group of experimental findings indicating that perceivers' judgments of the metric properties (e.g., length and surface orientation) of objects are quite inaccurate under many conditions (see, e.g., Norman, Todd, and Phillips [1995]; Norman *et al.* [1996]).[26] The fact that metric perception is so inaccurate has led some researchers to hypothesize that perhaps the visual system is primarily in

[26] Norman, Todd, and Phillips ([1995]) have found that judgments of surface orientation are inaccurate by an average of 14.5° even when subjects are given very reliable depth cues (e.g., binocular disparity, shading, texture, and motion). And Norman *et al.* ([1996]) found that subjects are highly inaccurate when asked to compare the lengths of line segments presented at random orientations in depth, though they are fairly accurate when asked to compare the lengths of nearby parallel lines. In particular, perceived length in depth tends to become progressively more compressed as a function of viewing distance, while length in the frontoparallel plane does not undergo this distortion. Incidentally, the finding that subjects are fairly accurate in comparing the lengths of parallel lines, but not non-parallel lines, lends some support to the view that perceivers represent affine shape. This is because the relative lengths of parallel line segments are preserved under affine transformations while the relative lengths of nonparallel line segments are not.

the business of producing estimates of more abstract shape properties—namely, affine properties. But how can this hypothesis be tested? One way is to place observers in restricted conditions in which *only* affine structure can be extracted (at least initially), and see whether it is indeed extracted.[27]

Researchers have explored this possibility quite extensively in the structure-from-motion paradigm.[28] In this paradigm, the observer is shown a display of dots or line segments that, when viewed statically, looks like a random 2-D configuration. The elements of the configuration, however, are generated by orthographic projection from elements of a (real or computer-generated) 3-D object. When the elements of the pattern begin to move in a way consistent with the movement of the 3-D object from which they were projected, the viewer spontaneously undergoes a percept of 3-D structure.

How does this happen? Ullman ([1979]) proved that it is possible to recover metric structure from a rigidly moving 3-D object on the basis of three distinct views (under orthographic projection) of four non-coplanar points of the object. For several years after Ullman's proof, it was assumed that the visual system solves the structure-from-motion problem by analyzing three views of the object and thus extracting precise metric shape.

However, while three views are necessary (and sufficient, assuming rigid movement) for extracting *metric* structure, it has been shown that with only two orthographic views, it is possible to recover the structure of an object modulo a uniform stretching in depth (Todd and Bressan [1990]; Ullman [1983]). In other words, one can recover the X and Y coordinates of each object point, but Z coordinates can only be recovered up to multiplication by a constant but unknown stretch factor k. As such, two views are sufficient to specify *relative* depth and *ratios* of distances along the depth axis, but are insufficient to specify *absolute* distances.

---

[27] There is another type of evidence that some have marshaled in support of affine shape perception. In a number of studies, Jan Koenderink and colleagues (e.g., Koenderink *et al.* [1996]; Koenderink *et al.* [2001]) have systematically investigated subjects' perceptual judgments of surface orientation. They have found that while perceivers' metric judgments are quite inaccurate (see note 27), perceived surface geometry nevertheless tends to be affine equivalent with real surface geometry. Specifically, perceived surface geometry tends to correspond to real surface geometry modulo a stretching or shearing in depth.

However, while some have taken these results to indicate that the visual system represents the affine properties of surfaces, I believe that this conclusion is too hasty. Rather, these findings can also be explained on the view that subjects *only* visually represent metric shape properties, but such representations simply tend to be non-veridical in systematic ways. That is, the visual system might non-veridically represent metric shapes that are distinct from, but affine equivalent to, metric shapes in the environment.

[28] For more—and critical—discussion of the evidence in favor of affine shape perception in structure-from-motion, see Bennett ([2012]).

Recall that affine properties are, roughly, those that are preserved under stretching or shearing along an arbitrary direction. These transformations preserve ratios of distances along parallel lines, but disrupt absolute distances. Accordingly, the types of geometrical properties recoverable on the basis of two views in a structure-from-motion display are, roughly (though not exactly), affine properties.[29] Thus, an interesting test case for the proposal that the visual system extracts affine shape is to present an observer with an apparent motion sequence involving just two orthographic projections of elements of a 3-D object (with other depth cues removed), and see whether this produces a percept of 3-D structure. If so, then a reasonable interpretation is that the visual system generated this percept by recovering the affine structure (more specifically, structure modulo an unknown stretch factor along the depth axis) specified by the two views. And indeed, a number of studies have suggested that subjects *do* undergo percepts of 3-D structure under these restricted conditions, and that they can accurately identify aspects of the affine structure of the object. For instance, subjects are able, on the basis of just two views, to discriminate curved from planar surfaces (Norman and Lappin [1992]), or determine whether two line segments are coplanar (Todd and Bressan [1990]). As such, it is reasonable to conclude that the visual system *can* recover affine shape properties.[30]

But are affine shape properties extracted in the *general* case, when more precise metric information *is* available (at least in principle)? There are theoretical reasons to think that they are. Critically, many affine properties/relations of line segments (collinearity, parallelism, straightness, and curvedness) are more readily computable on the basis of retinal images than metric properties. The reason is that they tend to be preserved under projection to the retina,[31] and moreover they are highly unlikely to arise at the retina by accident of viewpoint. As such,

[29] The type of structure recovered here is actually slightly more determinate than affine shape. Certain objects that are related by affine transformation *can* be distinguished on the basis of two orthographic views, if they differ by more than a uniform stretching in depth (e.g., a shear or a stretching along either the horizontal or vertical axis). See Todd and Bressan ([1990], p. 421).

[30] For now, I leave aside the issue of whether the structure-from-motion algorithms implemented by the visual system extract *only* affine shape. Todd and his colleagues have argued that the visual system is *incapable* of using more than two views in an apparent motion sequence to extract 3-D structure. This has, however, been challenged (Hogervorst and Eagle [2000]; Bennett *et al.* [2012]). Moreover, it is also possible that, on the basis of 2 views, subjects do perceive metric structure, but such percepts are generated on the basis of background heuristics rather than image data. However, even if this is right, it still seems plausible that such percepts are produced by way of prior representation of the affine structure determined by the 2 views (essentially, velocities of image elements).

[31] Parallelism is preserved under parallel or orthographic projection, but is not in general preserved under perspective projection. As such, it is not strictly speaking the case that parallelism is preserved under projection to the retina. However, when the ratio of an object's extension in depth to its distance from the viewer (known as the perspective ratio) is very small—as is often the case—the effects of perspective are negligible, and the projection process approximates parallel projection (see, e.g., Todd [1995]).

they are often called *nonaccidental properties* (e.g., Biederman [1987]). For instance, two collinear line segments in the world will always (discounting noise) project to collinear segments on the retinal plane—and the probability that two non-collinear segments in the world will project to collinear segments on the retinal plane is vanishingly small (Albert and Hoffman [1995]). Similar remarks hold for the other properties listed above. As such, detection of such properties at the retina is sufficient for inferring that they are present in the world.[32] By contrast, since metric properties (e.g., lengths and angles) are not preserved under projection, the visual system must do an incredible amount of computational work to recover them.

6 Against Metric Views of Visual Shape Representation

The evidence discussed in section 5 indicates that information about abstract shape properties is likely extracted by early visual processes and used in a number of ways. I now argue that this raises a serious difficulty for metric views of shape representation at the level of subpersonal visual processing.

The challenge for the metric view is simply to explain *how it is* that information about particular abstract shape properties is brought to bear in visual processing if visual shape representations do not make such information explicit. For as we saw earlier, 2½-D sketch-style (and other "view-based") representations make explicit only metric information, usually pertaining to individual surface points and edges (e.g., their numerical coordinates in a viewer-centered reference frame). Given just a representation of this sort, how can the visual system make use of the information that a certain object is, e.g., a triangle or a solid figure?

To make the problem more concrete, let's consider again the role that perception of topological properties plays in apparent motion perception.[33] Suppose that frame 1 contains a solid square, and that frame 2 contains both a square with a square-shaped hole and a solid

[32] This is not to suggest, however, that their detection at the retina is trivial. The detection of luminance edges—let alone geometrical relations between them—is an incredibly difficult computational task that has yet to be completely solved. The point, rather, is that such properties are in general *more easily* computable than metric features like depth and surface orientation.

[33] The emphasis on topology here is deliberate. Affine structure (or at least relations of affine equivalence across images) is often relatively easy to compute on the basis of image coordinates (see, e.g., Ullman [1996], pp. 208-13). As such, the representation of affine structure perhaps need not require drastic revisions to extant view-based models. Topology, however, is another story. It is well-known that topological properties (e.g., connectedness) are quite difficult to extract because (as can be proved) such properties in general cannot be computed by any set of local procedures that each depend only on a fixed set of points (Minsky and Papert [1969], pp. 12-14; Todd [2005]). As such, the perception of topology may require large changes to current models of shape processing (see also Chen [2005]).

triangle. If shown these two frames in succession, the subject is likely to see motion from the solid square to the solid triangle. And moreover, we have good reason to believe that it is sameness of topology (i.e., the property of being a solid figure) that accounts for this tendency. But for the visual system to compute motion paths in a manner that is selectively sensitive to topological similarities, respects of topological similarity must be *selected* or *highlighted* in contrast to other respects of geometrical similarity. That is, the respect in which the solid square is more similar to the solid triangle (both are solid figures) must be selected over the respects in which the solid square is more similar to the square with a hole (both have square-shaped bounding contours). The problem is that metric representations don't do this. Rather, in the 2½-D sketch, for instance, information about topology is *implicit alongside* information about, e.g., the lengths and angles of a shape's bounding contour. As such, this type of representation alone cannot provide the basis for mapping the solid square to the solid triangle rather than the square with a hole.

But what is it to *select* information about a certain abstract shape property? Plausibly, it is just to construct a representation that explicitly encodes the property. As such, it seems likely that visual shape representations explicitly encode information about abstract shape properties, such as topological properties and affine shape properties.[34]

This weighs in favor of a *layered* view of visual shape representation. On this approach, visual representations of geometrical properties are layered in a hierarchy roughly in accordance with the stability of those properties. Thus, when you see a triangular surface of an object, your visual system constructs *numerous* representations arranged in a multi-level hierarchy: The object is represented at one level as having a quite specific metric shape (e.g., a surface composed of points such-and-such a distance away with such-and-such orientation relative to the line of sight), but it is also represented at another level as a triangle, and at a third level as a solid (filled) figure. The explicit representation of such abstract properties can enable them to exert an influence on other visual or vision-based processes, such as motion perception, object tracking, and shape discrimination.

The hierarchical aspect of the proposal is critical. When the visual system represents an object as both square and quadrilateral, these two representations are almost certainly more

[34] Or, failing this, visual shape representations must at least explicitly encode *relations* of affine or topological equivalence among objects.

functionally integrated with one another than either is with, e.g., representing the object as maroon. Plausibly, the former two representations will need to be deployed in many of the same computational processes. Thus, representations of various shape properties should be appropriately related, and their relation arguably should reflect the asymmetric entailments among the properties represented. Hierarchical structure is the appropriate framework for accomplishing this (figure 5). Directed edges linking property representations at distinct levels of the hierarchy encode entailment relations between those properties.

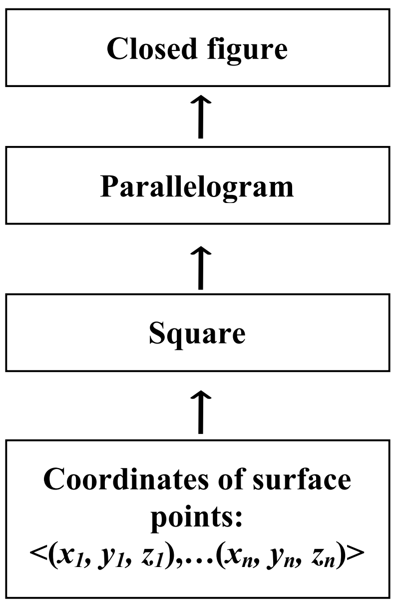

Figure 5. Hierarchy in accordance with geometrical stability. At the lowest level, highly unstable location features are represented. At the next level the property of being a square (invariant under similarity transformations) is represented. Next, the property of being a parallelogram (invariant under affine transformations). Finally, the property of being a closed figure (invariant under topological transformations).

Before moving on, I should forestall some potential misconceptions.

First, I should emphasize that I am *not* claiming that the visual system fails to construct a representation of metric features, such as the 2½-D sketch (though I remain agnostic about whether this is the best way to represent metric structure). Purely metric changes among affine equivalent objects are certainly registered by the visual system, and can be used to discriminate objects.[35] Indeed, if the visual system did not represent anything more specific than affine shape, then, as Li *et al.* ([2013]) note, a pizza box and a shoebox would be visually indistinguishable by shape. I am only claiming here that metric representations cannot *exhaust* the visual representation of shape. Shape representations must *also* explicitly encode information about more abstract shape properties.

[35] Lee *et al.* ([2012]) have shown that a novel target object can be distinguished fairly reliably from a metrically distinct (but affine equivalent) object so long as the viewer sees the target from a variety of perspectives.

Moreover, I am not claiming anything about the *order* in which the visual system extracts geometrical information on the basis of retinal input. Thus, it is possible that metric properties are extracted *prior* to abstract shape properties. This issue lies beyond my scope here. (However, there is reason to think that this is not the actual order of processing—see Chen [2005].)

Finally, the layered view should not be confused with the proposal that metric shape is represented in vision only at coarser levels of *precision* (which is almost certainly correct). For, just as an imprecise representation of being a poodle does not constitute a representation of being a dog, an imprecise representation of the sides and angular measurements of a particular triangle does not constitute a representation of the abstract property of being triangular.

7 Neural Underpinnings of Abstract Shape Perception

Given the strong psychophysical and theoretical support for visual representations of abstract shape, a natural next step is to inquire into their neural underpinnings. Vision scientists have started to take this step, and have so far met with promising results (for a more comprehensive review of this research, see Biederman [2013]).

Studies of both humans and nonhuman primates have produced compelling evidence that higher-level ventral stream neurons are more sensitive to changes in abstract shape than to mere metric changes. Thus, Kayaert, Biederman, and Vogels ([2003]) recorded the responses of single neurons in the anterior inferotemporal cortex (IT) of rhesus monkeys as they were shown a base stimulus, followed by a set of variations of the base stimulus. The variations included pairs of changes equated in their low-level image differences from the base, such that one change involved a variation in affine shape (e.g., a change from straight sides to curved, or from parallel to nonparallel), while the other involved mere metric change (e.g., a change in aspect ratio or degree of curvature).[36] In approximately 65% of the cases studied, neural responses were altered significantly more by a difference in affine shape than by an equated difference in metric properties (see also Vogels *et al.* [2001] and Kayaert *et al.* [2005]).

As regards the perception of topology, a recent fMRI study of the lateral occipital cortex (likely the human homologue of IT) has documented increased sensitivity to changes that disrupt the topological structure of a display (e.g., attaching two figures that were previously unattached)

[36] As noted above, "equating" low-level image differences is nontrivial. Kayaert et al. took the Euclidean distances between the gray levels of each pixel in the base stimulus and that pixel's counterpart in the variant stimulus.

in comparison to equated changes that do not disrupt topology (Kim and Biederman [2012]). Another fMRI study has revealed that when right-handed subjects perform a shape discrimination task, they exhibit greater activation in a region of the left inferior temporal gyrus when the figures to be discriminated are topologically distinct than when they are topologically equivalent (Wang *et al.* [2007]), suggesting that there may be some lateralization in the ventral stream processing of topological properties.

Thus, the existing neurophysiological data corroborates psychophysical findings, indicating increased sensitivity to changes in abstract shape, at least in certain cortical regions. The evidence also implicates ventral stream areas already known to be involved in visual shape processing (see Denys *et al.* [2004]).

8 Implications

Kulvicki ([2007]) has recently introduced the notion of *vertical articulateness* and argued that it is a highly general characteristic of perceptual content. On Kulvicki's characterization, a state has vertically articulate content "when for some property *P* that it represents, it also represents some *Q*, which is an abstraction from *P*" (Kulvicki [2007], p. 359). And roughly, one property is an abstraction from another only if there is an asymmetric entailment between the two: "*Q* is an abstraction from *P* only if being *P* entails being *Q* but the converse fails" (Kulvicki [2007], p. 359). While Kulvicki suggests that this view holds for a wide variety of types of perceptual content (e.g. color, shape, texture, etc.), he primarily motivates it in the case of color experience. Here I have mounted a sustained defense of the view that the content of visual states vis-à-vis geometrical properties is vertically articulate as well. This holds both for subpersonal representational states of the visual system, and for states of visual experience.

Kulvicki points out that vertically articulate perceptual content may have a crucial role to play in guiding concept acquisition. Most of our concepts concern fairly general categories. Thus, while most of us have the concept of red, few (if any) of us have concepts for maximally determinate shades of red. If perception presents us with the property redness (assuming there is such a property) in addition to presenting us with maximally determinate shades of red, then we have a much clearer picture of how we might acquire the concept of this general category. For, otherwise, generalization across specific shades would be left entirely up to post-perceptual cognition. Thus, it is quite possible that perceptual content *needs* to be vertically articulate if it is to form an adequate basis for learning.

Arguably, the need for vertical articulateness is even more pressing in the case of shape perception than in the case of color. It is well known that many of the earliest concepts children acquire reside at the so-called "basic" level, where perceptual "similarity" among members of the category is most salient (see Rosch [1978]). At this level, cars may be grouped together and distinguished from buses, but no general concept MOTOR VEHICLE is yet available. But how is such "similarity" to be characterized? A number of authors have contended that *shape* serves as the most important respect of perceptual resemblance during concept learning (see Rosch *et al.* [1976]; Landau *et al.* [1988]; Margolis [1998]).[37]

Nevertheless, while it is generally accepted that children must be sensitive to shape similarities during concept learning, members of the same basic category almost never share a common *metric* shape—e.g., some cars are longer or wider than others (and often drastically so). Rather, to locate the pertinent respect of shape similarity, arguably we must turn to abstract shape properties. And indeed, it appears that, at least in many cases, members of the same basic-level category (e.g., bottles or bowls) are at least roughly affine equivalent: the shape of one member can roughly be obtained from the shape of another by some combination of scaling, stretching, and shearing (see, e.g., Ons and Wagemans [2011]).

But how are children sensitive to similarities in abstract shape during the early stages of concept acquisition? Metric views of shape representation seem to have a difficult time answering this question, since abstract shape properties are left implicit in visual representation.[38] On the view outlined here, vision takes over much of the work that would otherwise have been left to cognition. That is, generalization across metrically distinct individuals occurs within vision. As such, on the layered view of visual shape perception I have offered, we gain a clearer conception of how visual shape perception may furnish a partial basis for early concept learning.

In closing, we should reject the idea that the representation of abstract shape properties belongs solely to the domain of post-perceptual cognition. Such properties are represented during

[37] However, it should be stressed that common shape is merely taken as a *guide* to common category membership. When information about "hidden" or "internal" features is available to children, it will often override shape information when making category judgments (see Gelman and Wellman [1991]).
[38] This difficulty is not new. View-based proposals have often been criticized for lacking a good account of basic level categorization (e.g., Palmer [1999], p. 452).

vision proper, exert an influence on other perceptual processes, and are represented in visual experience.

Acknowledgements

Thanks to Brian McLaughlin, Susanna Schellenberg, Andy Egan, Ben Bronner, and members of the Rutgers philosophy of mind research group for helpful discussion and feedback on earlier drafts of this paper. Thanks also to two anonymous reviewers, whose comments led to significant improvements. I am especially indebted to David Bennett, whose detailed comments led to substantial revisions of section 4. I also benefited from presenting parts of this material in Brian McLaughlin and Susanna Schellenberg's graduate seminar on spatial perception, at the 2014 Rutgers-Lund graduate philosophy conference, and as a poster at the 2014 annual meeting of the Society for Philosophy and Psychology.

References

Albert, M. and Hoffman, D. [1995]: 'Genericity in spatial vision', in D. Luce, K. Romney, D. Hoffman, and M. D'Zmura (*eds*), *Geometric Representations of Perceptual Phenomena: Articles in Honor of Tarow Indow on his 70th Birthday*, New York: Erlbaum, pp. 95-112.

Amir, O., Biederman, I., Herald, S.B., Shah, M.P., and Mintz, T.H. [2014]: 'Greater sensitivity to nonaccidental than metric shape properties in preschool children', *Vision Research*, **97**, pp. 83-8.

Barenholtz, E. and Tarr, M.J. [2008]: 'Visual judgment of similarity across shape transformations: Evidence for a compositional model of articulated objects', *Acta Psychologica*, **128**, pp. 331-8.

Belhumeur, P.N., Kriegman, D.J., and Yuille, A.L. [1999]: 'The bas-relief ambiguity', *International Journal of Computer Vision*, **35**, pp. 33-44.

Bennett, D.J. [2012]: 'Seeing shape: Shape appearances and shape constancy', *British Journal for the Philosophy of Science*, **63**, pp. 487-518.

Bennett, D.J., Zhao, H., Vuong, Q.C. and Liu, Z. [2012]: 'Humans can use information beyond 2 frames in structure from motion', Poster presented at VSS 2012.

Berkeley, G. [1710/1982]: *A Treatise Concerning the Principles of Human Knowledge*, Indianapolis, IN: Hackett Publishing Company.

Biederman, I. [1987]: 'Recognition-by-components: A theory of human image understanding', *Psychological Review*, **94**, pp. 115-47.

Biederman, I. [2013]: 'Psychophysical and neural correlates of the phenomenology of shape', in L. Albertazzi (*ed.*), *Handbook of Experimental Phenomenology: Visual Perception of Shape, Space, and Appearance*, Malden, MA: Wiley-Blackwell, pp. 417-36.

Biederman, I. and Bar, M. [1999]: ‘One-shot viewpoint invariance in matching novel objects’, *Vision Research*, **39**, pp. 2885-9.

Block, N. [2014]: ‘Seeing-as in the light of vision science’, *Philosophy and Phenomenological Research*, DOI: 10.1111/phpr.12135.

Brannan, D.A., Esplen, M.F., and Gray, J.J. [2012]: *Geometry*, Cambridge: Cambridge University Press (2nd edition).

Chen, L. [1982]: ‘Topological structure in visual perception’, *Science*, **218**, pp. 699-700.

Chen, L. [1985]: ‘Topological structure in the perception of apparent motion’, *Perception*, **14**, pp. 197-208.

Chen, L. [1990]: ‘Holes and wholes: A reply to Rubin and Kanwisher’, *Perception & Psychophysics*, **47**, pp. 47-53.

Chen, L. [2005]: ‘The topological approach to perceptual organization’, *Visual Cognition*, **12**, pp. 553-637.

Cole, M.S., Sanik, K., DeCarlo, D., Finkelstein, A., Funkhouser, T., Rusinkiewicz, S., & Singh, M. [2009]: ‘How well do line drawings depict shape?’, *ACM Transactions on Graphics*, **28**, pp. 1-9.

Denys, K., Vanduffel, W., Fize, D., Nelissen, K., Peuskens, H., Van Essen, D., and Orban, G.A. [2004]: ‘The processing of visual shape in the cerebral cortex of human and nonhuman primates: A functional magnetic resonance imaging study’, *The Journal of Neuroscience*, **24**, pp. 2551-65.

Edelman, S. [1997]: ‘Computational theories of object recognition’, *Trends in Cognitive Sciences*, **1**, pp. 296-304.

Edelman, S. [1999]: *Representation and Recognition in Vision*, Cambridge, MA: MIT Press.

Evans, G. [1985]: ‘Molyneux’s question’, in A. Phillips (*ed.*), *Gareth Evans: Collected Papers*, Oxford: Clarendon Press. Reprinted in A. Noë & E. Thompson (*eds*), *Vision and Mind: Selected Readings in the Philosophy of Perception*, Cambridge, MA: MIT Press, pp. 319-49.

Feldman, J. [2007]: ‘Formation of visual “objects” in the early computation of spatial relations’, *Perception & Psychophysics*, **69**, pp. 816-27.

Fodor, J.A. [1983]: *Modularity of Mind*, Cambridge, MA: MIT Press.

Gelman, S.A. and Wellman, H.M. [1991]: ‘Insides and essences: Early understandings of the non-obvious’, *Cognition*, **38**, pp. 213-44.

Graf, M. [2006]: ‘Coordinate transformations in object recognition’, *Psychological Bulletin*, **132**, pp. 920-45.

Green, M. [1986]: ‘What determines correspondence strength in apparent motion?’, *Vision Research*, **26**, pp. 599-607.

Hoffman, D.D. [1998]: *Visual Intelligence: How We Create What We See*, New York: Norton.

Hogervorst, M.A. and Eagle, R.A. [2000]: ‘The role of perspective effects and accelerations in perceived three-dimensional structure-from-motion’, *Journal of Experimental Psychology: Human Perception and Performance*, **26**, pp. 934-55.

Hummel, J.E. and Stankiewicz, B.J. [1996]: ‘Categorical relations in shape perception’, *Spatial Vision*, **10**, pp. 201-36.

Jackendoff, R. [1987]: *Consciousness and the Computational Mind*, Cambridge, MA: MIT Press.

Kayaert, G., Biederman, I., and Vogels, R. [2003]: ‘Shape tuning in macaque inferior temporal cortex’, *The Journal of Neuroscience*, **23**, pp. 3016-27.

Kayaert, G., Biederman, I., and Vogels, R. [2005]: ‘Representation of regular and irregular shapes in macaque inferotemporal cortex’, *Cerebral Cortex*, **15**, pp. 1308-21.

Kayaert, G. and Wagemans, J. [2010]: ‘Infants and toddlers show enlarged visual sensitivity to nonaccidental compared with metric shape changes’, *i-Perception*, **1**, pp. 149-58.

Kellman, P.J. [2003]: ‘Visual perception of objects and boundaries: A four-dimensional approach’, in R. Kimchi, M. Behrmann, and C.R. Olson (*eds*), *Perceptual Organization in Vision: Behavioral and Neural Perspectives*, Mahwah, NJ: Lawrence Erlbaum, pp. 155-201.

Kim, J.G., & Biederman, I. [2012]: ‘Greater sensitivity to nonaccidental than metric changes in the relations between simple shapes in the lateral occipital cortex’, *NeuroImage*, **63**, 1818-26.

Kirsh, D. [2003]: ‘Implicit and explicit representation’, in L. Nadel (*ed.*), *Encyclopedia of Cognitive Science*, Vol. 2, London: Macmillan, pp. 478-81.

Koenderink, J.J., van Doorn A.J. and Kappers A.M.L. [1996]: ‘Pictorial surface attitude and local depth comparisons’, *Perception & Psychophysics*, **58**, pp. 163-73.

Koenderink, J.J., van Doorn, A.J., Kappers, A.M.L. and Todd, J.T. [2001]: ‘Ambiguity and the ‘mental eye’ in pictorial relief’, *Perception*, **30**, pp. 431-48.

Kulvicki, J. [2007]: ‘Perceptual content is vertically articulate’, *American Philosophical*

Quarterly, **44**, pp. 357-69.

Landau, B., Smith, L. B. and Jones, S. S. [1988]: 'The importance of shape in early lexical learning', *Cognitive Development*, **3**, pp. 299-321.

Lee, Y-L., Lind, M., Bingham, N. and Bingham, G.P. [2012]: 'Object recognition using metric shape', *Vision Research*, **69**, pp. 23-31.

Li, Y., Sawada, T., Shi, Y., Steinman, R.M. and Pizlo, Z. [2013]: 'Symmetry is the *sine qua non* of shape', in S.J. Dickinson and Z. Pizlo (*eds*), *Shape Perception in Human and Computer Vision*, New York: Springer, pp. 21-40.

Lu, Z.-L. and Sperling, G. [2001]: 'Three-systems theory of human visual motion perception: Review and update', *Journal of the Optical Society of America A*, **18**, pp. 2331-70.

Margolis, E. [1998]: 'How to acquire a concept', *Mind & Language*, **13**, pp. 347-69.

Marr, D. [1982]: *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*, San Francisco: W.H. Freeman.

Marr, D. and 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**, pp. 269–294.

Minsky, M. and Papert, S.A. [1969]: *Perceptrons*, Cambridge, MA: MIT Press.

Nakayama, K. and Shimojo, S. [1992]: 'Experiencing and perceiving visual surfaces', *Science*, **257**, pp. 1357-63.

Norman, J.F. and Lappin, J.S. [1992]: 'The detection of surface curvatures defined by optical motion', *Perception & Psychophysics*, **51**, pp. 386-96.

Norman, J.F., Todd, J.T. and Phillips, F. [1995]: 'The perception of surface orientation from multiple sources of optical information', *Perception & Psychophysics*, **57**, pp. 629-36.

Norman, J.F., Todd, J.T., Perotti, V.J. and Tittle, J.S. [1996]: 'The visual perception of three-dimensional length', *Journal of Experimental Psychology: Human Perception and Performance*, **22**, pp. 173-86.

Ons, B. and Wagemans, J. [2011]: 'Development of differential sensitivity for shape changes resulting from linear and non-linear planar transformations', *i-Perception*, **2**, pp. 121-36.

Palmer, S.E. [1999]: *Vision Science: Photons to Phenomenology*, Cambridge, MA: MIT Press.

Palmer, S.E. [2003]: 'Perceptual organization and grouping', in R. Kimchi, M. Behrmann, and C.R. Olson (*eds*), *Perceptual Organization in Vision: Behavioral and Neural*

Perspectives, Mahwah, NJ: Lawrence Erlbaum, pp. 3-43.

Peacocke, C. [1992]: *A Study of Concepts*, Cambridge, MA: MIT Press.

Phillips, F., Todd, J.T., Koenderink, J.J., and Kappers, A.M. [2003]: 'Perceptual representation of visible surfaces', *Perception & Psychophysics*, **65**, pp. 747-62.

Pinker, S. [1997]: *How the Mind Works*, New York: W.W. Norton & Company.

Pylyshyn, Z.W. [1999]: 'Is vision continuous with cognition? The case for cognitive impenetrability of visual perception', *Behavioral and Brain Sciences*, **22**, pp. 341-423.

Riesenhuber, M. and Poggio, T. [2002]: 'Neural mechanisms of object recognition', *Current Opinion in Neurobiology*, **12**, pp. 162-8.

Rosch, E.H. [1978]: 'Principles of categorization', in E. Rosch and B. Lloyd (*eds*), *Cognition and Categorization*, Hillsdale, NJ: Erlbaum, pp. 27-48.

Rosch, E.H., Mervis, C.B., Gray, W.D., Johnson, D.M. and Boyes-Braem, P. [1976]: 'Basic objects in natural categories', *Cognitive Psychology*, **8**, pp. 382-439.

Sekuler, A. and Palmer, S. [1992]: 'Perception of partly occluded objects: A microgenetic analysis', *Journal of Experimental Psychology: General*, **121**, pp. 95-111.

Siegel, S. [2010]: *The Contents of Visual Experience*, Oxford: Oxford University Press.

Schwitzgebel, E. [2011]: *Perplexities of Consciousness*, Cambridge, MA: MIT Press.

Tarr, M.J. and Pinker, S. [1989]: 'Mental rotation and orientation-dependence in shape recognition', *Cognitive Psychology*, **21**, pp. 233-82.

Todd, J.T. [1995]: 'The visual perception of 3-dimensional structure from motion', in W. Epstein and J. Rogers (*eds*), *Handbook of Perception and Cognition, Volume 5: Perception of Space and Motion*, Orlando, FL: Academic Press, pp. 201-26.

Todd, J.T. [2004]: 'The visual perception of 3D shape', *Trends in Cognitive Sciences*, **8**, pp. 115-21.

Todd, J.T. [2005]: 'Stability and change', *Visual Cognition*, **12**, pp. 639-42.

Todd, J.T. and Bressan, P. [1990]: 'The perception of 3-dimensional affine structure from minimal apparent motion sequences', *Perception & Psychophysics*, **50**, pp. 509-23.

Todd, J.T., Chen, L. and Norman, J.F. [1998]: 'On the relative salience of Euclidean, affine, and topological structure for 3D form discrimination', *Perception*, **27**, pp. 273-82.

Todd, J.T., Weismantel, E. and Kallie, C.S. [2014]: 'On the relative detectability of configural properties', *Journal of Vision*, **14**, pp. 1-8.

Tse, P.U. [1999]: 'Volume completion', *Cognitive Psychology*, **39**, pp. 37-68.

Ullman, S. [1979]: 'The interpretation of structure from motion', *Proceedings of the Royal Society of London. Series B. Biological Sciences*, **203**, pp. 405-26.

Ullman, S. [1983]: 'Recent computational studies in the interpretation of structure from motion', in J. Beck, B. Hope, and A. Rosenfeld (*eds*), *Human and Machine Vision*, New York: Academic Press, pp. 459-80.

Ullman, S. [1996]: *High-Level Vision: Object Recognition and Visual Cognition*, Cambridge, MA: MIT Press.

Ullman, S. [1998]: 'Three-dimensional object recognition based on the combination of views', *Cognition*, **67**, pp. 21-44.

Ullman, S. and Basri, R. [1991]: 'Recognition by linear combinations of models', *IEEE Transactions on Pattern Analysis & Machine Intelligence*, **13**, pp. 992-1006.

Veltkamp, R.C. and Latecki, L.J. [2006]: 'Properties and performance of shape similarity Measures', in *Proceedings of the 10th IFCS Conference on Data Science and Classification*, Slovenia, July 2006.

Vogels, R., Biederman, I., Bar, M., and Lorincz, A. [2001]: 'Inferior temporal neurons show greater sensitivity to nonaccidental than to metric shape differences', *Journal of Cognitive Neuroscience*, **13**, pp. 444-53.

Wang, B., Zhou, T.G., Zhuo, Y., and Chen, L. [2007]: 'Global topological dominance in the left hemisphere', *Proceedings of the National Academy of Sciences*, **104**, pp. 21014-9.

Wagemans, J., Elder, J.H., Kubovy, M., Palmer, S.E., Peterson, M.A., Singh, M., and von der Heydt, R. [2012]: 'A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure-ground organization', *Psychological Bulletin*, **138**, pp. 1172-217.

Williams, P. and Tarr, M.J. [1999]: 'Orientation-specific possibility priming for novel three-dimensional objects', *Perception & Psychophysics*, **61**, pp. 963-76.

Zhou, K., Luo, H., Zhou, T., Zhuo, Y. and Chen, L. [2010]: 'Topological change disturbs object continuity in attentive tracking', *Proceedings of the National Academy of Sciences*, **107**, pp. 21920-4.