Structure Constancy

1. Introduction

Imagine that you are meeting a friend for coffee, and you see her walking toward your table. As she walks, her arms and legs turn about their joints. Moreover, her forearms turn slightly about her elbows, and her tibias move about her knees. I suggest that, despite these changes, her *overall structure* seems to remain *stable*. Call this phenomenon *structure constancy*. Structure constancy is ubiquitous in our visual experiences of objects. In this paper I'll offer an account of structure constancy, and then I'll argue that the phenomenon has important consequences for viable theories of the subpersonal underpinnings of visual spatial phenomenology.

I'll begin in section 2 with a general discussion of perceptual constancy, and then I'll identify an important respect in which structure constancy differs from the more familiar geometrical constancies. In section 3, I'll offer a characterization of *compositional structure*, and argue that structure constancy involves experientially representing an object as retaining compositional structure across certain geometrical changes. In section 4, I'll argue that the phenomenology of structure constancy cannot be underpinned by a representational format that fails to make part structure explicit, and that this has implications for identifying the locus of visual shape phenomenology within visual system processing. In section 5, I'll argue that structure constancy raises a problem for views on which the visual representation that underlies our experience of spatial/geometrical properties is wholly viewer-centered. I suggest that our visual experience of geometrical properties plausibly reflects the simultaneous deployment of multiple reference frames for specifying location.

2. Perceptual Constancy

2.1. What is perceptual constancy?

Most theorists agree that perceptual constancy involves a type of *stability* in one's perceptual response across certain *changes* (cf. Cohen forthcoming). Thus, Tyler Burge (2010) writes: "Perceptual constancies are capacities systematically to represent a particular or an attribute as the same despite significant variations in registration of proximal stimulation" (408). Similarly, Stephen Palmer characterizes (visual) perceptual constancy as "the ability to perceive the properties of environmental objects, which are largely constant over different viewing conditions, rather than the properties of their projected retinal images, which vary greatly with viewing conditions" (Palmer 1999: 125).[1]

Under these characterizations, to display perceptual constancy with respect to a property *P*, one must at minimum perceptually represent *P* across changes in the way one's sensory organs are stimulated. In the case of vision, this would be to perceptually represent *P* across changes in the stimulation of retinal cells.[2]

While Burge's definition provides a useful starting point, it has a significant drawback. Burge does not say what is involved in representing a particular or attribute "as the same" across variations in proximal stimulation. On one reading, this would require that a subject (or a perceptual system) represent that something perceived under one condition of proximal stimulation *is the same*—or, at least, the same in respect of a particular attribute, such as color—as something perceived under a different condition of proximal stimulation. On another reading,

[1] Roughly this notion appears in the work of a variety of authors, such as Michaels and Carello (1981: 20), Rock (1983: 24), Smith (2002), and Pizlo (2008). Other notions of constancy instead focus on the stability of one's perceptual representation across changes in a property's *appearance* (e.g., Shoemaker 2000; Noë 2004; Hill 2014).

[2] Notice that these changes may be either *intra*-object or *inter*-object: a perceiver may continue to perceive the *same* object as *P* despite changes in the proximal stimulation received from it, or a perceiver may perceive *different* objects as both being *P* despite differences in the proximal stimulation received from them.

it would require only that one perceptually attribute the same property *P* to individuals encountered under different conditions of proximal stimulation.

The first notion is more demanding. To represent that things perceived under different conditions are the same in respect of a particular property, one must be able to perceptually represent *comparisons* or *relations* between those things. This might involve either representing that some property *P* is shared by things perceived in different conditions, or retained by a single thing perceived in different conditions. There is no such requirement in order to simply represent the same property *P* under two different conditions. We can call the first notion the *strong* type of constancy, and the latter the *weak* type. I'll suggest below that structure constancy is generally of the strong type.

2.2. Geometrical constancies

To set up the rest of the paper, I want to briefly apply this account to geometrical constancies in particular. I'll understand an object's "geometrical properties" to include its size, shape, and location. Moreover, I'll henceforth focus on the strong type of perceptual constancy, where one not only recovers a property under two different conditions, but also represents that the property is shared or retained across changes in proximal stimulation.

To delineate the nature of geometrical constancy under this characterization, we need to know what features—or "cues"—within proximal stimulation are relevant to recovering distal geometrical properties. Research indicates that in the case of shape and size perception, there are a number of such cues—e.g., 2-D retinal shape, context, shading, texture, and motion, among others (Palmer 1999: ch. 5). For the sake of simplicity, however, let's just focus on 2-D retinal shape. Accordingly, our paradigm case of geometrical constancy in what follows will be one in which a subject perceptually represents an object as retaining a geometrical property (distal shape or size) across changes in the shape or size of its retinal projection.

Changes in the shape or size of an object's retinal projection result from *transformations* of the object within a retinocentric frame of reference (a frame of reference built around an origin and intrinsic axes of the retina). For instance, if the object undergoes a *rotation* transformation where it is slanted in depth relative to the line of sight, this issues in a change in the shape of its projection on the retina. A circular object presents a circular image when seen straight on, but an elliptical image when seen at a slant.

Geometers classify transformations as *rigid* or *non-rigid*. Rigid transformations are ones that don't involve any changes to an object's intrinsic metric properties. By the "metric properties" of an object, I have in mind, roughly, those properties of the object that depend essentially on its constituent edge lengths, angles, and curvature. For instance, a metric property of a square surface is the property of having four angles of 90°. Rigid transformations include translation (simple change of position), rotation, and reflection (change in "handedness"). Such transformations do not alter the distances or angles between points of the transformed object.

Non-rigid transformations, on the other hand, do involve changes to an object's intrinsic metric properties. The simplest kind of non-rigid transformation is uniform scaling, in which an object changes in size but its angles stay the same. Other non-rigid transformations include stretching, shearing, skewing, and bending, which disrupt both lengths and angles.[3] Both rigid and non-rigid transformations usually result in changes to an object's 2-D retinal shape. For

[3] In geometry, transformations are arranged into groups, such as affine transformations, projective transformations, and topological transformations. The group consisting only of rigid transformations and uniform scaling is the *similarity group*. Often, the similarity group is taken to be definitive of what we mean when we say that two objects have the "same shape." We mean that one can be brought into precise register with the other by some composition of similarity transformations (Palmer 1999: 364-365).

example, if a square is stretched into an oblong rectangle, this will usually be associated with a change in the shape of its projection on the retina.

Size constancy involves seeing things as sharing/retaining a property across rigid transformations in a retinocentric reference frame (since non-rigid transformations usually change an object's size). For example, one might perceptually represent something as retaining a particular size property despite viewing it at different distances. Shape constancy, as it is normally introduced, involves seeing things as sharing/retaining a property despite either a rigid transformation (e.g., rotation or translation with respect to the retina and the line of sight), or uniform scaling. For example, one might see an object as retaining a particular distal shape despite viewing it at different orientations (slants) in depth.

Structure constancy cannot be reduced to size or shape constancy. The reason is that structure constancy involves seeing an object as retaining a property (which I'll label "compositional structure") across certain non-rigid transformations that (unlike uniform scaling) disrupt both the distances and angles among parts of the object. As such, structure constancy is distinctive insofar as the transformations relevant to exercising structure constancy are different from (and, as we'll see, more geometrically complicated than) the transformations relevant to exercising the other geometrical constancies.

3. The Visual Phenomenology of Structure Constancy

Many of the most ecologically significant objects with which we interact are *biological objects*—especially animals and other humans. Many biological objects have an important characteristic: When they move, they *change shape*. This happens when, for instance, a person walks across a room. Even though the person's precise metric properties are constantly changing, intuitively we are able to see her body as retaining some important aspects of structure as she moves. In this section I will first introduce the notion of *compositional structure*. Then I will propose that structure constancy is most plausibly explained by the proposal that visual experience represents compositional structure.

3.1. Compositional structure introduced

Objects often seem to decompose naturally into parts. For example, the object in figure 1a seems to have three natural parts, as shown in figure 1b.

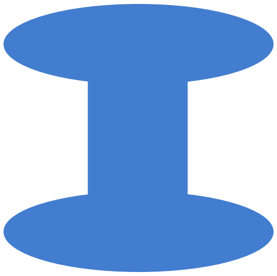


Figure 1a

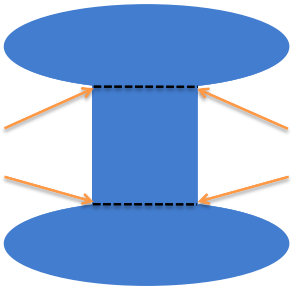


Figure 1b

In addition to being intuitively compelling, judgments about an object's decomposition into parts are remarkably consistent across observers (e.g., De Winter & Wagemans 2006). This, in addition to its role in several well-known theories of object recognition (Marr & Nishihara 1978; Biederman 1987), has led part decomposition to become a topic of extensive research in perceptual psychology.[4]

Critically, there are *rules* by which the visual system parses objects into parts. An important rule for our purposes is called the *minima rule*, first formulated by Hoffman and Richards (1984). The minima rule states that the boundaries between the perceived parts of an object tend to be found at extrema of negative curvature—roughly, places at which the surface of the object is locally most concave. Concave regions are, intuitively, regions where the object's surface curves "inward." Figures 2a and 2b illustrate applications of the minima rule in specifying part boundaries.

[4] For general discussions, see Singh and Hoffman (2001), and Hoffman (2001).

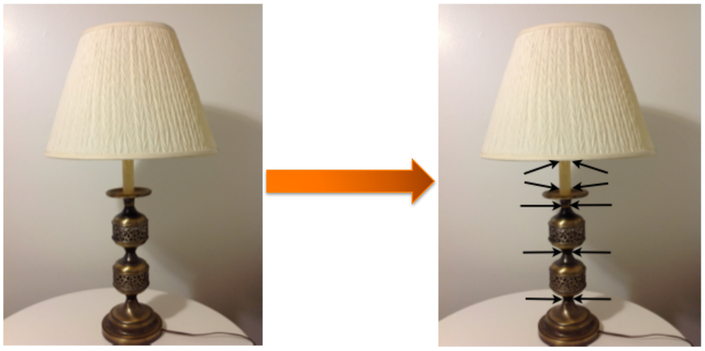


Figure 2a

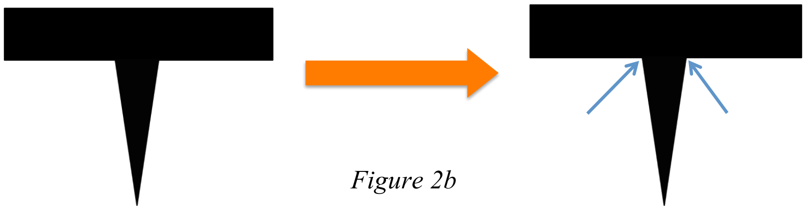


Figure 2b

While the minima rule tells us where to find boundaries between parts, it does not tell us precisely how to “slice” an object. That is, it does not specify how to make part *cuts*. Fortunately, this problem has also been studied extensively. Other things being equal, part cuts tend to obey the *short-cut rule* (Singh, Seyranian, & Hoffman 1999), which states that the visual system prefers part cuts that link negative minima of curvature, and generally opts for the shortest such links possible. The part cuts in figure 1b conform to the short-cut rule, as would the most obvious cuts of figures 2a and 2b.[5]

The representation of part decomposition (roughly in accordance with the minima and short-cut rules) has incredible psychological utility (e.g., Ling & Jacobs 2007). For example, many objects that move non-rigidly nevertheless change shape in a systematic manner. Roughly, their parts retain their intrinsic shapes, though the spatial relations between parts may change. The moving human body, as we saw, is an instance of this generalization, but so are the moving bodies of most other animals, along with many manufactured devices (such as, e.g., a stapler or a

[5] However, these rules have exceptions. See Singh and Hoffman (2001) for discussion.

reclining chair). By decomposing a complex object into parts one can predict the ways it is likely to transform over time. It is disposed to move in ways that alter the spatial relations between parts, but unlikely to move in ways that either alter the intrinsic shapes of parts or displace the joints about which the parts rotate.

We are now ready to introduce the notion of compositional structure. A compositional structure of an object *O* consists of the following:

1. A decomposition of *O* into a pairwise disjoint set of (proper) parts $P_1 ... P_n$,
2. The approximate part-centered locations of boundaries between connected pairs of parts in $P_1 ... P_n$,
3. The approximate intrinsic shapes of $P_1 ... P_n$.

Structure constancy amounts, I suggest, to the ability to perceptually represent an object as retaining a particular compositional structure across proximal cue variations (e.g., changes in retinal shape) that result from non-rigid transformations of the object.

A terminological note: A set of parts $P_1 ... P_n$ will be called "pairwise disjoint" if and only if for all pairs (P_i, P_j) drawn from $P_1 ... P_n$, P_i and P_j do not overlap. Now, three substantive remarks on the visual representation of compositional structure:

First, according to my characterization of compositional structure, an object will have at least as many compositional structures as it has decompositions into parts. This may give rise to some initial concerns. For decompositions are *cheap*. An object can be decomposed in any number of ways, and it certainly does not seem as though we perceptually experience *all* of these decompositions, much less perceive them all as remaining stable as an object moves. However, the explanation of structure constancy offered here does not rely on this claim. Rather, the idea is that a *particular* compositional structure is perceptually represented, while the others are not.

Second, note that I take only the *approximate* intrinsic shapes of parts to figure in compositional structure. Due to, say, the deformation of muscle tissue, a person's upper arm does not retain its metric properties precisely as the arm rotates. So it is likely that to perceive an object as retaining compositional structure over time, the object's parts need only retain their shapes up to some more coarse-grained standards of precision.

Third, note that the locations of part boundaries must be specified in part-centered reference frames. This means that the locations of part boundaries are represented via their spatial relations to certain points on the connected parts themselves. The reason is this: If, say, the location of a perceived person's elbow (a boundary between forearm and upper arm) is specified in a viewer-centered reference frame, then its location *does* change as the person moves. Similarly, if its location is specified in a simple object-centered reference frame (e.g., with an origin at the center of gravity of the person's body), then its location changes as a result of rotation of the upper arm about the shoulder joint. Only when the elbow's location is specified in a frame of reference centered on either the forearm or upper arm (according to their intrinsic axes) does its location remain approximately stable across nonrigid movement of the body. Like the representation of metric part shapes, the representation of part boundaries should be somewhat coarse-grained. Even in a part-centered reference frame, part boundaries do not remain *perfectly* stable across non-rigid movement.

3.2. Representing compositional structure in experience

I've proposed that the compositional structure of an object is represented in visual experience, and that this is what accounts for the experience of structure constancy. But this claim requires further defense. In what follows I'll defend it using a modification of Susanna Siegel's method of phenomenal contrast (Siegel 2010).

Siegel's method is introduced as a procedure for determining whether visual experiences represent a given property *F*. It requires us to examine two overall experiences that differ phenomenally, and determine whether the best explanation of their phenomenal contrast is that one of the overall experiences contains a visual experience that represents *F*, while the other does not.

Unfortunately, Siegel's method of phenomenal contrast cannot be straightforwardly applied in the current case. Consider any two experiences *A* and *B* that phenomenally differ, and are plausible candidates for differing vis-à-vis the compositional structures they represent. The method asks us to determine whether the phenomenal contrast between *A* and *B* is *best* explained by the hypothesis that they indeed differ with respect to the visual experiential representation of compositional structure. However, for any two such experiences, there will plausibly be *numerous other* differences in their visual experiential content, and some of these other differences would also seem to plausibly explain the phenomenal contrast.

Notice that if an object loses a particular compositional structure, it must cease to occupy precisely the same spatial region. For example, any change in the intrinsic shape of an object *O*'s part *P* necessitates a change in *O*'s compositional structure, but it also necessitates a change in the precise spatial region that *O* occupies. Thus, if we only consider this individual change, the difference in phenomenology that accompanies successive experiences of *O* (before and after the change) may seem to be explained just as well by the hypothesis that visual experience only represents the precise spatial region that *O* occupies, rather than *O*'s compositional structure.

How should we evaluate the hypothesis that visual experiences represent compositional structure? I suggest that, rather than examining two *individual* experiences, we ought to examine *pairs of changes* in experience. We begin with an experience of a *base stimulus*, and a hypothesis about the particular compositional structure *C* of the base stimulus represented in experience. Next, we consider the experiences of two *test stimuli*. Test stimulus 1 shares compositional structure *C* with the base stimulus, while test stimulus 2 does not. However, *both* test stimuli differ from the base stimulus in their precise metric structure. If visual experiences represent compositional structure, then one might expect the difference between one's experiences of the base stimulus and test stimulus 2 to be *more salient* than the difference between one's experiences of the base stimulus and test stimulus 1.

However, for this to be a fair test, we need to ensure that, as regards factors *besides* compositional structure, the change from the base stimulus to test stimulus 1 is either roughly comparable to, or else greater than, the change from the base stimulus to test stimulus 2. In particular, we want to ensure that the increase in salience accompanying the change between the base stimulus and test stimulus 2 is not due to a greater difference in local features of the stimuli, or to a greater "overlap" in their spatial regions.

There are a variety of ways to measure the amount of local point or feature difference between two figures (see, e.g., Kayaert et al. 2003; Veltkamp & Latecki 2006). Perhaps the most straightforward measure is "Hamming distance" (Ullman 1996: 5). To find this distance, we first specify the two figures within a coordinate system. Each is represented by a binary vector indicating, for each point *p* within the coordinate system, whether *p* "belongs" to the figure ("1" if it belongs, "0" if it does not). Given this, we measure the distance between the two figures by normalizing the figures to a standard position and orientation, then summing the places in which the vectors for the two figures differ.

In what follows I'll only consider cases in which the Hamming distance between the base stimulus and test stimulus 1 is clearly either greater than, or roughly comparable to, the Hamming distance between the base stimulus and test stimulus 2. The argument is that if the difference between the base and test stimulus 2 is more phenomenologically salient under these conditions, then the *best explanation* is that visual experience represents the base stimulus as sharing a property with test stimulus 1, but doesn't attribute this property to test stimulus 2. My proposal is that the former two are visually experienced as sharing a compositional structure.

Consider figures 3a-3c. Let 3a serve as our base stimulus. Its compositional structure *C* plausibly consists of the following: a decomposition into head, torso, arms, and legs; the approximate intrinsic shapes of these parts; and the joints at which they are connected to one another. Figure 3b (test stimulus 1) shares *C* with the base stimulus. Figure 3c (test stimulus 2) does *not* share *C* with the base stimulus (joint locations are changed). Phenomenologically, I find these changes to be qualitatively different. The transformation to 3b seems "natural," while the transformation to 3c doesn't, even though the Hamming distance between the base and test stimulus 1 is obviously greater than the distance between the base and test stimulus 2. The proposal that visual experience represents compositional structure explains this. In the first case, the two objects are visually experienced as sharing a feature (a particular compositional structure), while in the second case, they are not.

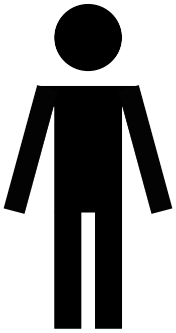


Figure 3a

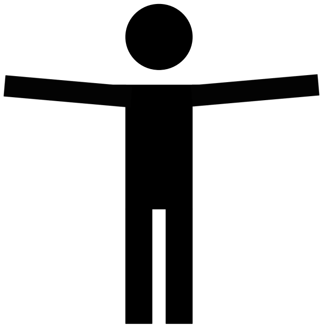


Figure 3b

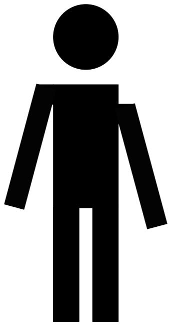
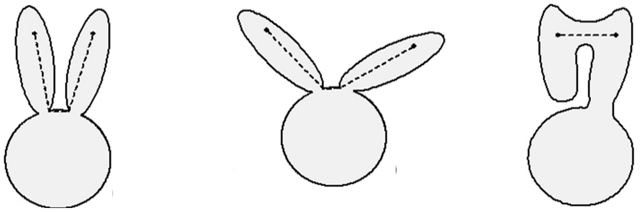


Figure 3c

Consider another example, due to Ling and Jacobs (2007). The base stimulus is shown in figure 4a, while the test stimuli 1 and 2 are shown in figures 4b and 4c, respectively. Again, the Hamming distance between the base and test stimulus 1 is greater (i.e., the two have less overlap in local features), but the transition between the two arguably seems less salient (and also more natural) than the transition from the base to test stimulus 2. Once again, test stimulus 1 preserves compositional structure, while test stimulus 2 does not.



Figures 4a-4c (left to right). Source: Ling & Jacobs (2007)

3.3. A post-perceptual explanation?

There are two potential worries with examples involving human bodies, bunny ears, and the like. First, it is unclear whether the contrast in salience here is due to *visual* experience, or rather to postperceptual expectations given familiarity with such objects and the ways they move. Second, even if the example does reveal the representation of compositional structure in visual experience, it is unclear how general its implications are. Perhaps compositional structure is represented in visual experience only for highly familiar figures, and not for decomposable figures in general. For these reasons, it would be more persuasive if such contrasts in salience could be demonstrated using novel shapes.

Evidence suggests that compositional structure is extracted for novel shapes. Barenholtz and Tarr (2008) showed subjects a novel base shape, along with two transformations of the base shape. Only one of these transformations—which I'll again label test stimulus 1—preserved compositional structure under the minima and short-cut rules. The shape that failed to preserve compositional structure—test stimulus 2—could involve either a change in location of a boundary between parts, or a change in a part's intrinsic shape. Figure 5 shows a case in which test stimulus 2 involves a change of the former type. The differences between the base stimulus and test stimuli 1 and 2 are essentially equated in their low-level feature changes, because the narrower part on the right of the figure was rotated the same amount in both cases. The only difference was whether the part's axis of rotation was its joint with the rest of the object (preserving compositional structure) or its endpoint (altering compositional structure).

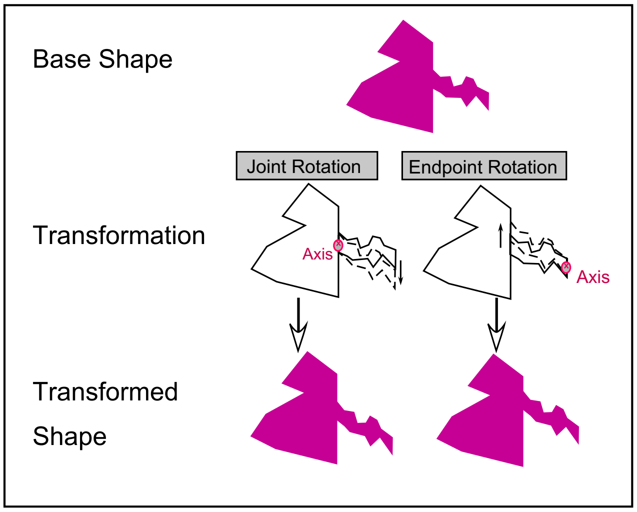


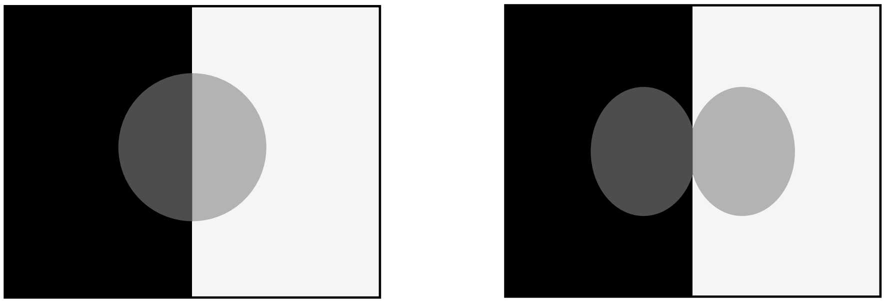
Figure 5. The transformed shape at bottom left (test stimulus 1) preserves the compositional structure of the base (intrinsic part shapes, and locations of part boundaries). The transformed shape at bottom right (test stimulus 2) alters compositional structure, because the part boundary shifts upward. Source: Barenholtz & Tarr (2008).

Participants saw the three shapes, and were simply asked to indicate which of the transformed shapes was more similar to the base. Barenholtz and Tarr found that subjects were significantly more likely to indicate that the shape that preserved compositional structure was more similar. The same pattern of results was obtained with other triples of shapes where the change that disrupted compositional structure instead altered the intrinsic shape of the base stimulus's part, rather than its joint location. Thus, there is evidence that the ability to extract compositional structure is highly general and not limited to particular classes of familiar objects (e.g., human or animal shapes).

Nevertheless, how do we know that compositional structure isn't recovered post-perceptually, even in the case of novel objects? If this were the case, then structure constancy wouldn’t really deserve to be labeled a *perceptual* constancy at all. Although it is difficult to settle the matter conclusively, there is empirical evidence that the representation of compositional structure is genuinely perceptual.

I’ve held that the explanation of structure constancy resides in the visual system’s ability to decompose objects into parts and represent their boundaries and shapes independently. Because of this, visual experience distinguishes transformations that preserve a given compositional structure from those that do not. As such, evidence for the perceptual differentiation of *parts* provides support for the view that structure constancy is perceptual.

Perhaps the strongest evidence that parts are extracted perceptually is that part decomposition influences other paradigmatically perceptual processes. A striking example of this involves the perception of transparency. Compare figures 6a and 6b.



Figures 6a (left) and 6b (right). Source: Singh & Hoffman (2001)

While figure 6a appears to depict a transparent gray filter in front of a half-dark, half-light background, in figure 6b the percept of transparency is greatly diminished (Singh & Hoffman 1998). Rather, the occluding object is perceived as an opaque figure with two differently shaded regions. The received explanation for this is that the visual system expects regions of a single part of an object to have the same reflectance, but it does not expect regions of different parts of an object to have the same reflectance (or at least it expects this less strongly). Since the object in figure 6b can be broken down into two natural parts, it can be interpreted as an opaque figure whose parts have different reflectances.[6]

If part decomposition interacts with other perceptual processes, we have strong evidence that it is a perceptual process as well. For, while it is possible to advert to a cognitive penetration account in these cases, I can think of no motivation for doing so (aside from a pretheoretical conviction that part decomposition *must* be cognitive). Moreover, it is worth noting that the tendency to parse objects into parts also strikes me as involuntary—I cannot *help* seeing many objects as decomposed into natural parts. This is another hallmark feature of a perceptual process (e.g., Fodor 1983; Pylyshyn 1999).[7]

Even if parts are represented during perception, this does not yet show conclusively that *compositional structure* is represented during perception. The representation of compositional structure involves both decomposing an object into parts *and* (i) representing the intrinsic shape of each part independently, and (ii) representing the part-centered boundaries between parts.

With respect to (i), there are good reasons to believe that the visual system encodes the shapes of different parts separately from one another, and independently of their spatial relations. Though this hypothesis was initially put forth on computational and theoretical grounds (e.g., Biederman 1987; Marr & Nishihara 1978; Palmer 1978), there is now compelling experimental evidence for it. Consider a recent study of the subject S.M., an individual with integrative agnosia. Integrative agnosia is a visual disorder that affects processes involving the integration of local visual information into a global percept. Behrmann et al. (2006) found that S.M. was capable of correctly discriminating sequentially presented objects from one another when the

[6] Part perception has also been argued *inter alia* to influence figure-ground organization (Hoffman & Singh 1997), the spread of visual attention (Barenholtz & Feldman 2003) and pop-out effects in visual search (Xu & Singh 2002).
[7] Further evidence for the automaticity of part decomposition is provided by studies of human infants. Using a dishabituation paradigm, Bhatt et al. (2010) have provided compelling evidence that 6 ½ month-old infants are sensitive to the minima and short-cut rules.

objects differed in the intrinsic shape of a single part (e.g., a cube-shaped part versus an ellipsoid-shaped part), but, unlike "normal" participants, he could not discriminate objects when they differed purely in their parts' spatial configuration (e.g., a cube to the left of a cylinder versus a cube on top of a cylinder). In line with (i), this suggests that there are visual processes that extract the shapes of individual parts, and these processes can remain intact despite an inability to extract the global configuration of an object (see also Davidoff & Roberson 2002; Cacciamani, Ayars, & Peterson 2014). Again, absent defeating evidence, I conclude that the processing of individual part shapes happens within perception.

The claim, (ii), that part boundaries are represented in part-centered reference frames is the hardest to establish. Before covering empirical support for this claim, we need to get clearer on what part-centered reference frames are, and how they have been developed in the vision science literature.

Constructing a reference frame involves choosing a set of parameters so that the position of any point is uniquely determined by specifying its values on these parameters (Klatzky 1998). When a reference frame is *centered* on an object *O*, this means that the positions of points are encoded at least partly in terms of their spatial relations (e.g., distance and direction) to a point, or set of points, on *O*. For example, to construct a polar coordinate system, we first stipulate an origin *o* and an axis *A* through *o*, and then specify the location of any given point *p* in terms of two parameters: its distance from *o*, and the angle between *A* and the line from *o* to *p*.

Many shape representation theorists have proposed that the visual system recovers, roughly, the *medial axis structure* of an object (e.g., Blum & Nagel 1978; Rosenfeld 1986; Kimia 2003; Feldman & Singh 2006). The medial axis of a figure is composed of the set of points having two or more closest points on the bounding contour of the figure. A figure's medial axis generally looks like a "skeleton" from which the figure is "grown." These schemes are centered on the points that compose the axis. Roughly, they represent the positions of points on the boundary of the shape by specifying their distances and directions from corresponding points on the axis.

Importantly, in a wide range of cases, the medial axis structure of an object bears a close relation to its decomposition into parts under the minima and short-cut rules.[8] This is because different parts of the object tend to be associated with distinct axis branches (see figure 7). Thus, if the visual system extracts the medial axis structures of objects, and distinct parts are associated with distinct axis branches, then these distinct axis branches can be used to construct separate reference frames each centered on a distinct part. Accordingly, evidence for the visual representation of medial axis structure also counts as evidence that the visual system uses part-centered frames of reference.

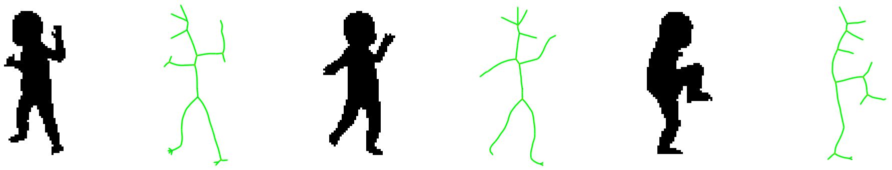


Figure 7. The medial axis structure of three human silhouettes. Note that in most cases the intuitive parts (arms, torso, and legs) correspond to distinct axis branches. Source: Kimia (2003).

The prediction that vision extracts medial axis structure has recently been confirmed using a very simple paradigm. Firestone and Scholl (2014) showed subjects a novel shape, asked them to tap the shape wherever they liked, and recorded the locations of subjects' taps. If

[8] In practice, however, the correspondence is not perfect. In standard models (e.g., Blum & Nagel 1978), small perturbations of a shape's contour give rise to "spurious" axis branches that do not intuitively correspond to distinct parts of the shape. Feldman and Singh (2006) have recently developed a novel, Bayesian approach to axial description that "cleans up" the medial axis representation. The axes returned by their model are not medial axes, although for smooth shapes without many perturbations their axes closely resemble medial axes. Feldman and Singh's model builds in a prior favoring smoother axes with fewer branches, and its results tend to better match intuitive part cuts.

“tapping” behavior is guided by a visual shape representation that specifies the intrinsic (perhaps medial) axes of object parts, one might expect the locations of subjects’ taps to be influenced by these axes. Sure enough, Firestone and Scholl found that the recorded taps (when aggregated) corresponded closely to the medial axes of the shapes presented. That is, subjects were much more likely to tap an object somewhere along its medial axis than they were to tap other regions of the shape. This provides compelling evidence that medial axis structure is automatically extracted by vision, since the task did not require subjects to attempt to extract these axes.

If the visual system represents spatial properties and relations by using intrinsic part axes, then it should encode the parts of an object as retaining their spatial relations to one another across transformations in viewer-centered position and orientation. For as long as these transformations are rigid, the *part-centered* relations between the constituents of the configuration will not change.

There is intriguing evidence that areas of the visual system code for medial axis structure independently of viewer-centered position. In a recent fMRI study, Lescroart and Biederman (2013) presented subjects with figures that differed in either their medial axis configuration, their component part shapes, or their viewer-centered orientation. Figure 8 displays some of these figures: Shapes in the same row share the same medial axis structure, though their orientations and intrinsic part shapes vary. Stimuli in the same column share the same intrinsic part shapes, but differ in medial axis structure. The line segments next to the figures indicate viewer-centered orientation. Lescroart and Biederman found that by area V3, patterns of BOLD activity could classify stimuli according to shared medial axis structure at a rate significantly better than chance (even though such stimuli differed in the shapes of their component parts), and classification of medial axis structure was significantly more accurate than classification of orientation (whereas the opposite pattern was observed in V1). This provides at least *prima facie* evidence that extrastriate areas of the visual system represent configurations according to spatial arrangements of part axes.

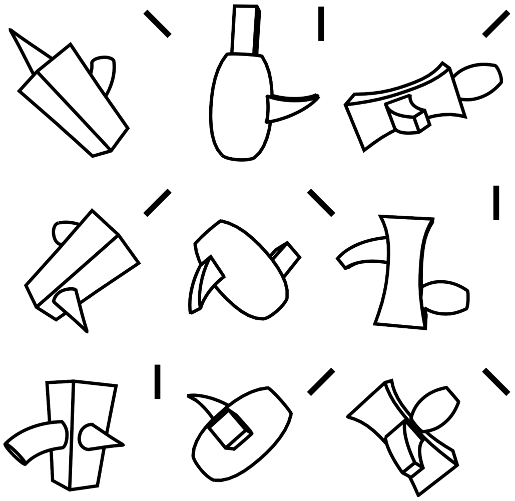


Figure 8. Stimuli used by Lescroart and Biederman (2013). Source: Lescroart and Biederman (2013).

We have encountered strong evidence that each of the characteristics that figure in compositional structure is recovered during vision proper. I contend that the proposal that compositional structure is represented in subpersonal visual processing and manifests itself in visual experience offers the best account in light of all the evidence at our disposal, including both the patterns of phenomenological salience associated with shape transformations in objects, and the empirical data on subpersonal shape processing.

4. Mereological Structure and Shape Representation Schemes

What does structure constancy tell us about the subpersonal underpinnings of shape experience? I believe it has at least two important consequences for these underpinnings. In this section, I'll argue that structure constancy has the consequence that certain aspects of shape experience must be underpinned by a representation scheme that is *mereologically structured*. In the next, I'll argue that because structure constancy must recruit non-viewer-centered reference frames, it raises problems for recent approaches on which spatial phenomenology is subserved by some enrichment of Marr's 2½-D sketch.

Call a representation *R* mereologically structured iff:

(1) *R* purports to introduce individuals *O* and *O** independently, and
(2) *R* represents that *O* is a proper part of *O**.

To purport to introduce *n* individuals *independently* is to deploy *n* distinct representational items that each purport to introduce distinct individuals.[9] For instance, the phrases "John's cat" and "John's dog" purport to introduce two individuals independently. The central idea, then, is that if a representation *R* is mereologically structured, then distinct constituents of *R* purport to pick out distinct entities that are related through mereological composition, and *R* represents the parthood relations that those entities stand in to one another.

Some shape representation schemes are not mereologically structured. Consider, for instance, schemes found in the *view-based* approach to object recognition (see, e.g., Ullman and Basri 1991; Ullman 1996; Edelman 1999; Riesenhuber & Poggio 2002). On several of these models, the representation of shape just amounts to the representation of a vector composed of the viewer-centered feature coordinates of some of the object's "critical features"—e.g., vertices, inflection points, and curvature maxima.[10] This type of scheme does not incorporate the representation of parthood at all, and proponents of the view-based approach have often downplayed the role of part decomposition in visual processing (e.g., Edelman 1999: 89-94).

[9] The notion of introducing an individual is left deliberately vague. For present purposes, it does not matter whether individuals are introduced in vision by description or by singular reference. But for recent defenses of the latter view, see Pylyshyn (2007) and Recanati (2012).

[10] Such views have been offered primarily in order to account for findings indicating that object recognition is sensitive to viewpoint. Such results have sometimes been believed problematic for hierarchical approaches to shape representation, which generally invoke non-viewer-centered reference frames. However, for an argument that hierarchical models can accommodate viewpoint effects on recognition, see Bar (2001).

Perhaps the most popular mereologically structured scheme is *hierarchical description* (see, e.g., Palmer 1977; Marr & Nishihara 1978; Feldman 2003; Leek et al. 2009; Hummel 2013). A hierarchical description (figure 9) is a representational structure that contains distinct nodes corresponding to each individual introduced, encodes either mereological or spatial relations between nodes, and associates monadic featural information with each node. It is usually depicted as a tree.

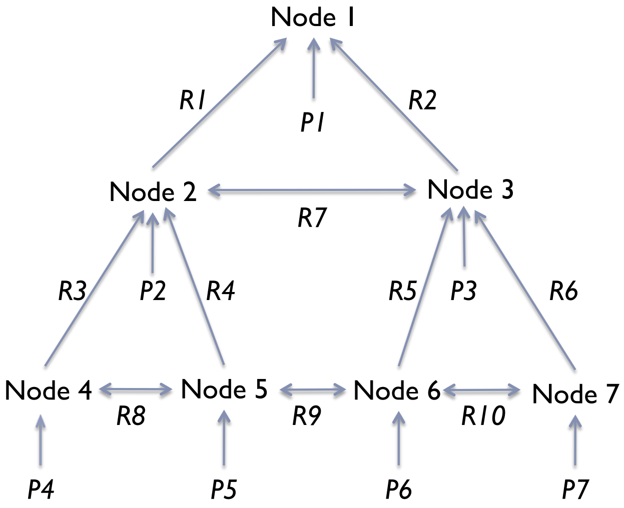


Figure 9. The format of a hierarchical description.

Edges traversing levels of a hierarchical description represent parthood. For present purposes, I'll assume that the visual system's representation of parthood is transitive: If a hierarchical description represents O_1 as part of O_2 and O_2 as part of O_3, then it also represents O_1 as part of O_3. Edges linking nodes at the same level of a description represent spatial relations between parts. Let's call edges representing parthood *P-edges*, and edges representing spatial relations *S-edges*. A subset of the S-edges will describe the locations of *boundaries* between parts: They will represent, for a pair of connecting parts, the points where those parts meet (in part-centered coordinates). Call these *B-edges*.

Several models of shape processing invoke both an earlier, view-based stage and a later, hierarchical stage (Marr 1982; Hummel 2001, 2013). If this is right, it is natural to ask which (if either) of these stages underpins shape phenomenology. I argue that structure constancy provides strong reason to locate at least certain aspects of shape phenomenology at the hierarchical stage.

It is hard to see how a view-based scheme could underpin structure constancy. Because view-based schemes do not introduce the parts of objects as distinct individuals, such models do not *prioritize* any particular part decomposition over others. Each of the many possible decompositions of an object into parts is compatible with, say, the same arrangement of vertices and curvature extrema along the object's bounding contour. Because view-based schemes fail to prioritize a specific part decomposition, they lack the resources for distinguishing changes that leave intrinsic part shapes intact while altering the global shape of the object from changes that alter the intrinsic shapes of parts. Indeed, any given change could—relative to *some* decomposition—be considered a change in the intrinsic shapes of an object's parts. Thus, without a specification of which decomposition is the *relevant* one, it is impossible to decide whether a particular change does or does not deform intrinsic part shapes.

Hierarchical description, on the other hand, can be applied to the explanation of structure constancy. Let's spell this out using the human body as an example. Given a hierarchical description that introduces a human body *O*, a *compositional structure* of *O* is encoded in (i) the intrinsic shape information associated with nodes at some level of the description lower than the level at which *O* is introduced, such as a level introducing the head, torso, arms, and legs, (ii) the P-edges linking these nodes to the node introducing *O*, and (iii) the B-edges linking these nodes to one another—e.g., torso-centered locations of the shoulders, where the arms intersect the torso. By distinguishing this information from the information encoded by the remaining S-edges (such as the angle formed between an arm and the torso) and information about *O*'s global metric structure, the representation enables the visual system to distinguish transformations that leave a given compositional structure intact from those that do not. As such, hierarchical descriptions may underpin structure constancy.

5. Comparison with Other Approaches

Many have been attracted to the idea that visual phenomenology seems to present us with an array of facing surfaces, rather than, e.g., the 2-D retinal image or the volumetric structure of objects (e.g., Jackendoff 1987; Tye 1991, 1995; Prinz 2012). In light of this, theorists influenced by Marr's (1982) pioneering tripartite theory of vision have sought to locate the underpinnings of visual consciousness at the "intermediate" level of processing, which describes the geometry of surfaces. In Marr's framework, the intermediate level is occupied by the 2½-D sketch, so theorists have often appealed to the 2½-D sketch, though usually with some alterations or enrichments, which I'll discuss below.

The 2½-D sketch 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: 275-279). It can be construed as a type of "depth map" representing certain spatial properties of thousands of very small surface patches within one's field of vision. The important thing to note is that the 2½-D sketch lacks two features that I have argued are central to explaining structure constancy. First, the scheme is mereologically unstructured. This is because the 2½-D sketch only attributes geometrical features to very small surface patches in one's field of vision, and it does not represent the composition of such surface patches into larger individuals. Second, the scheme is *wholly viewer-centered*. That is, all locations in the visual field are represented relative to an origin centered on the viewer. So the locations of part boundaries are not represented in part-centered coordinates.

Jackendoff (1987) calls on the 2½-D sketch in his account of the subpersonal underpinnings of conscious experience, but recognizes that Marr's representational structure has important defects (e.g., lack of explicit surface segmentation, perceptual grouping, etc.). As such, he develops an *enriched* 2½-D sketch, which he calls the 2½-D structural description (see Jackendoff 1987: 331-338). More recently, Prinz (2012) has appealed to Jackendoff's theory in his "intermediate view" of the subpersonal basis of visual consciousness.

Jackendoff enriches Marr's depth map with the primitive elements *boundary* and *region*, and the predicates *directed*, *abutting*, *overflow*, and *occlusion*. Boundaries and regions are obtained by appropriately segmenting the initially undifferentiated 2½-D sketch. The predicates represent properties and relations of these boundaries and regions. For example, the 2½-D structural description has the resources to encode (via the *directedness* predicate) figure-ground relations, and can encode (via the *overflow* predicate) that a region extends outside one's field of vision. Moreover, Jackendoff also incorporates parthood into his 2½-D structural description. He proposes that boundaries are identified not only where one finds luminance edges in the retinal image, but also in accordance with Hoffman and Richards' minima rule.

For our purposes, the important point is this. Jackendoff's model organizes the visual array into objects and parts, but it does not alter the basic reference frame of the 2½-D sketch. The depth map is segmented, and certain properties of segmented regions are represented, but the underlying coordinate frame remains wholly viewer-centered. Likewise, although Prinz (2012) offers some revisions to Jackendoff's model, he agrees that the representation underlying visual consciousness is wholly viewer-centered (Prinz 2012: 50-57). See also Tye (1991: 90-97; 1995: 140-141) for a similar view.

For the reasons canvassed above, viewer-centered representational schemes cannot plausibly underpin structure constancy. Whenever an object moves relative to the perceiver, the viewer-centered locations of its part boundaries change. But to explain the patterns of phenomenological salience associated with shape transformations, we need a representation that treats part boundaries as remaining *stable* across such changes in viewer-centered location, so long as they don't shift their positions relative to the connected parts themselves. A part-centered scheme does this, while a viewer-centered scheme does not.

As such, the view I have offered importantly departs from these approaches on a critical dimension of shape representation (viz., its reference frame), though it does have a feature in common with them (viz., incorporating part-based organization).

I should underscore, however, that the view that the locations of certain things are experienced in part-centered reference frames does not imply—or even suggest—that we *fail* to also experience things in a viewer-centered reference frame. Indeed, it is an undeniable aspect of our phenomenology that we perceive from a perspective, e.g., that objects are seen to have certain spatial relations to our point of view (e.g., Peacocke 1992; Schellenberg 2008; Bennett 2009). Nevertheless, I think that on the most plausible analysis, vision represents geometrical/spatial properties within *multiple* reference frames simultaneously (cf. Briscoe 2009; Humphreys et al. 2013). Indeed, the view that perception uses multiple reference frames seems to comport best with the *overall* phenomenology of watching a non-rigid object move. When a person walks, for example, there is a sense in which her joint locations seem to stay stationary, but also a sense in which they seem to move relative to your viewpoint.

6. Conclusion

In this paper I have offered an account of a novel type of geometrical constancy, which I've called *structure constancy*. I argued that we visually experience objects as retaining their compositional structure despite certain changes that alter their intrinsic metric properties. Moreover, I have drawn out implications of structure constancy for both the representational content and the subpersonal underpinnings of visual shape experience.

References

Bar, M. (2001). Viewpoint dependency in visual object recognition does not necessarily imply viewer-centered representation. *Journal of Cognitive Neuroscience*, *13*, 793-799.

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

Barenholtz, E., & Tarr, M.J. (2008). Visual judgment of similarity across shape transformations: Evidence for a compositional model of articulated objects. *Acta Psychologica*, *128*, 331-338.

Behrmann, M., Peterson, M.A., Moscovitch, M., & Suzuki, S. (2006). Independent representation of parts and the relations between them: Evidence from integrative agnosia. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(5), 1169-1184.

Bennett, D.J. (2009). Varieties of visual perspectives. *Philosophical Psychology*, *22*(3), 329-352.

Bhatt, R.S., Hayden, A., Kangas, A., Zieber, N., & Joseph, J.E. (2010). Part perception in infancy: Sensitivity to the short-cut rule. *Attention, Perception, & Psychophysics*, *72*(4), 1070-1078.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, *94*(2), 115-117.

Blum, H. & Nagel, R. N. (1978). Shape description using weighted symmetric axis features. *Pattern Recognition*, *10*(3), 167-180.

Briscoe, R. (2009). Egocentric spatial representation in action and perception. *Philosophy and Phenomenological Research*, *79*, 423-460.

Burge, T. (2010). *Origins of Objectivity*. Oxford: Oxford University Press.

Cacciamani, L., Ayars, A.A., & Peterson, M.A. (2014). Spatially rearranged object parts can facilitate perception of intact whole objects. *Frontiers in Psychology*, *5*, 1-11.

Cohen, J. (forthcoming). Perceptual constancy. To appear in M. Matthen (ed.), *Oxford Handbook of Philosophy of Perception*. Oxford: Oxford University Press.

Davidoff, J., & Roberson, D. (2002). Development of animal recognition: A difference between parts and wholes. *Journal of Experimental Child Psychology*, *81*, 217-234.

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

Edelman, S. (1997). Computational theories of object recognition. *Trends in Cognitive Sciences*, *1*, 296-304.

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

Feldman, J. (2003). What is a visual object? *Trends in Cognitive Sciences*, *7*(6), 252-256.

Feldman, J., & Singh, M. (2006). Bayesian estimation of the shape skeleton. *Proceedings of the National Academy of Sciences*, *103*(47), 18014-18019.

Feldman, J., Singh, M., Briscoe, E., Froyen, V., Kim, S., & Wilder, J. (2013). An integrated Bayesian approach to shape representation and perceptual organization. In S.J. Dickinson & Z. Pizlo (eds.), *Shape Perception in Human and Computer Vision*, pp. 55-70. New York: Springer.

Firestone, C., & Scholl, B.J. (2014). "Please tap the shape, anywhere you like": Shape skeletons in human vision revealed by an exceedingly simple measure. *Psychological Science*, DOI: 10.1177/0956797613507584.

Fodor, J.A. (1983). *Modularity of Mind*. Cambridge, MA: MIT Press.

Graf, M. (2006). Coordinate transformations in object recognition. *Psychological Bulletin*, *132*, 920-945.

Hill, C.S. (2014). The content of visual experience. In *Meaning, Mind, and Knowledge*, pp. 218-238. Oxford: Oxford University Press.

Hoffman, D.D. (2001). Mereology of visual form. In C. Arcelli, L.P. Cordella, & G.S. di Baja (eds.), *Visual Form 2001*, pp. 40-50. New York: Springer.

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

Hoffman, D.D., & Singh, M. (1997). Salience of visual parts. *Cognition*, *63*, 29-78.

Hummel, J. E. (2001). Complementary solutions to the binding problem in vision: Implications for shape perception and object recognition. *Visual Cognition*, *8*, 489-517.

Hummel, J.E. (2013). Object recognition. In D. Reisburg (ed.), *Oxford Handbook of Cognitive Psychology*, pp. 32-46. Oxford: Oxford University Press.

Humphreys, G., Gillebert, C.R., Chechlacz, M., & Riddoch, M.J. (2013). Reference frames in visual selection. *Annals of the New York Academy of Sciences*, *1296*, 75-87.

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

Kayaert, G., Biederman, I., & Vogels, R. (2003). Shape tuning in macaque inferior temporal cortex. *The Journal of Neuroscience*, *23*, 3016–27.

Kimia, B.B. (2003). On the role of medial geometry in human vision. *Journal of Physiology-Paris*, *97*, 155-190.

Klatzky, R.L. (1998). Allocentric and egocentric spatial representations: Definitions, distinctions, and interconnections. In C. Freksa, C. Habel, & K.F. Wender (eds.), *Spatial Cognition*, pp. 1-17. Berlin: Springer.

Leek, E.C., Reppa, I., Rodriguez, E., & Arguin, M. (2009). Surface but not volumetric part structure mediates three-dimensional shape representation: Evidence from part–whole priming. *The Quarterly Journal of Experimental Psychology*, *62*(4), 814-830.

Lescroart, M.D., & Biederman, I. (2013). Cortical representation of medial axis structure. *Cerebral Cortex*, *23*, 629-637.

Ling, H., & Jacobs, D. W. (2007). Shape classification using the inner-distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *29*(2), 286-299.

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

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

Michaels, C.F., & Carello, C. (1981). *Direct Perception*. Englewood Cliffs, NJ: Prentice-Hall.

Noë, A. (2004). *Action in Perception*. Cambridge, MA: MIT Press.

Palmer, S.E. (1977). Hierarchical Structure in Perceptual Representation. *Cognitive Psychology*, *9*, 441-474.

Palmer, S.E. (1978). Fundamental aspects of cognitive representation. In E. Rosch & B. Lloyd (eds.), *Cognition and Categorization*, pp. 261-304. Hillsdale, NJ: Lawrence Erlbaum.

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

Peacocke, C. (1992). *A Study of Concepts*. Cambridge, MA: MIT Press.

Pizlo, Z. (2008). *3D Shape: Its Unique Place in Visual Perception*. Cambridge, MA: MIT Press.

Prinz, J.J. (2012). *The Conscious Brain: How Attention Engenders Experience*. Oxford: Oxford University Press.

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

Pylyshyn, Z.W. (2007). *Things and Places: How the Mind Connects with the World*. Cambridge, MA: MIT Press.

Recanati (2012). *Mental Files*. Oxford: Oxford University Press.

Riesenhuber, M. & Poggio, T. (2002). Neural mechanisms of object recognition. *Current Opinion in Neurobiology*, *12*, 162-8.

Rock, I. (1983). *The Logic of Perception*. Cambridge, MA: MIT Press.

Rosenfeld, A. (1986). Axial representations of shape. *Computer Vision, Graphics, and Image Processing*, *33*(2), 156-173.

Schellenberg, S. (2008). The situation-dependency of perception. *The Journal of Philosophy*, *105*, 55-84.

Shoemaker, S. (2000). Introspection and phenomenal character. *Philosophical Topics*, *28*, 247-273.

Siegel, S. (2010). *The Contents of Visual Experience*. Oxford: Oxford University Press.

Singh, M. & Hoffman, D. D. (1998). Part boundaries alter the perception of transparency. *Psychological Science*, *9*, 370-378.

Singh, M., & Hoffman, D.D. (2001). Part-based Representations of Visual Shape and Implications for Visual Cognition. In T.F. Shipley & P.J. Kellman (eds.), *From Fragments to Objects: Segmentation and Grouping in Vision*. New York, NY: Elsevier Science.

Singh, M., Seyranian, G.D., & Hoffman, D.D. (1999). Parsing silhouettes: The short-cut rule. *Perception & Psychophysics*, *61*, 636-660.

Smith, A.D. (2002). *The Problem of Perception*. Cambridge, MA: Harvard University Press.

Tye, M. (1991). *The Imagery Debate*. Cambridge, MA: MIT Press.

Tye, M. (1995). *Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind*. Cambridge, MA: MIT Press.

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

Ullman, S. & Basri, R. (1991). Recognition by linear combinations of models. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, *13*, 992-1006.

Veltkamp, R.C. & 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.

Xu, Y., & Singh, M. (2002). Early computation of part structure: Evidence from visual search. *Perception & Psychophysics*, *64*(7), 1039-1054.