The puzzle of cross#modal shape experience

THE PUZZLE OF CROSS-MODAL SHAPE EXPERIENCE

E. J. Green

Massachusetts Institute of Technology

Abstract: The *puzzle of cross-modal shape experience* is the puzzle of reconciling the apparent differences between our visual and haptic experiences of shape with their apparent similarities. This paper proposes that we can resolve the cross-modal puzzle by reflecting on another puzzle. The *puzzle of perspectival character* challenges us to reconcile the variability of shape experience through shifts in perspective with its constancy. An attractive approach to the latter puzzle holds that shape experience is complex, involving both *perspectival* aspects and *constant* aspects. I argue here that parallel distinctions between perspectival and constant aspects of shape experience arise in sight and touch, and that perspectival aspects are modality-specific while at least some constant aspects are constitutively multisensory. I then address a powerful challenge to the idea that aspects of spatial phenomenology are shared cross-modally.

1. Introduction

Suppose you first see a pencil, and then you close your eyes and hold it in your hands. In both cases you perceptually experience the pencil's shape. But is your experience of its shape the same or different in the two cases?

Some say it is thoroughly different. Lopes (2000) writes: "It seems to me that tactile and visual experiences have distinctive phenomenal characters through and through. What it is like to see the shape of a cube is different from what it is like to touch the same shape" (445).[i] Others disagree. While visual and haptic experiences differ in myriad ways, perhaps the phenomenology of *shape* is nonetheless the same across modalities. O'Callaghan (2019) writes: "It is not obvious that a visual experience of, say,

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/nous.12384.

sphericity—when it is considered in abstraction from the experience of other visible features—must differ in respect of each phenomenal feature from a tactual experience of sphericity" (126).[ii]

The relation between visual and haptic shape experience has long captivated philosophers. It features prominently in debates about Molyneux's question (Locke, 1694/1975; Evans, 1985; Levin, 2008; Schwenkler, 2013; Matthen & Cohen 2020) and concept empiricism (Berkeley, 1709/1965; Prinz, 2002, ch. 5). More recently, cross-modal differences in shape phenomenology have figured in challenges to representationalist theories of experience, where it has been argued that visual and haptic shape experiences differ in phenomenal character without differing in content (Peacocke, 1983, 27-28; Lopes, 2000; O'Dea, 2006).

Visual and haptic experiences differ in many ways. Many of these differences aren't puzzling. Since we are aware of color in vision but not in touch, it isn't puzzling that we enjoy color phenomenology in vision but not touch. But we are also aware of some of the *same* properties in both modalities. We are aware of shape in both sight and touch, and our awareness of shape has a distinctive phenomenology in both cases that is arguably separate from the phenomenology of, say, color or temperature. The question, then, is whether this distinctive phenomenology of shape is the same or different across modalities.

Both answers have some intuitive appeal. When I compare my visual and haptic experiences of the pencil, it doesn't seem like they differ *solely* with respect to the phenomenology of color or temperature. When I focus on the pencil's shape specifically, vision and touch seem to afford different perspectives on the very same property, and this difference is manifest in experiential phenomenology. On the other hand, visual and haptic shape experiences also seem to have commonalities. Visually experiencing a cube is surely *more* like haptically experiencing a cube than haptically experiencing a sphere. If so, then visual and haptic shape phenomenology plausibly have some aspects in common—aspects that are shared in the former case but not the latter. Call the puzzle of reconciling the apparent differences between visual and haptic shape experiences with their apparent similarities the *puzzle of cross-modal shape experience*, or the *cross-modal puzzle* for short.

I propose that we can resolve the cross-modal puzzle by considering a different puzzle: the *puzzle of perspectival character*. When you look at a coin as it is rotated, the coin appears to remain the same shape throughout, but its appearance also changes. The challenge is to reconcile the apparent *variability* of shape experience with its apparent *constancy*. An attractive answer holds that the visual appearance of shape is multi-tiered. *Perspectival* elements of visual shape phenomenology change with shifts in perspective, while *constant* elements remain invariant. I claim that a parallel distinction between perspectival and constant elements arises in touch. Specifically, I distinguish three tiers of shape phenomenology, arguing that analogous tiers arise in both modalities. The first two tiers are perspectival, while the third is not. Moreover, the first two are modality-specific, while the third is constitutively multisensory. The solution to the cross-modal puzzle thus lies in the internal complexity of shape phenomenology.

2. Two Tempting Proposals

This section considers two tempting views about what distinguishes visual and haptic shape experience. Unfortunately, both face difficulties.

One thought is that visual and haptic experiences represent shape at different levels of grain or specificity.[iii] Perhaps haptic experience represents more determinable or coarse-grained geometrical features than visual experience does. This view would suggest a tidy solution to the cross-modal puzzle. Suppose that haptic experience represents an object as having 5-7 sides, while visual experience represents it as exactly 6-sided. Then the two experiences might be similar thanks to representing compatible values of the determinable *shape*, but distinct because visual experience represents a more determinate value of that determinable.

It turns out, however, that haptic shape perception can be remarkably precise. Blindfolded participants can haptically distinguish a cylinder with a circular cross-section from one whose cross-section is an ellipse with an aspect ratio of just 1.03 (van der Horst & Kappers, 2008). This approximates thresholds for visually distinguishing a circle from an ellipse (Zanker & Quenzer, 1999). Thus, it's doubtful that the difference between visual and haptic shape experience is due to systematically greater determinacy in vision.

Another approach holds that visual shape experience constitutively involves the experience of a more encompassing *space* that objects are seen to occupy, whereas haptic shape experience does not. Richardson (2014) puts the idea as follows:

> Visually appearing (or looking) square is appearing to extend into or take up space that we are aware of as a space in which things can be seen. The same is not true of tactually appearing (or feeling) square. The phenomenology of feeling square is not that of appearing to occupy a space in which things can be seen in the square-ish way, or in any way at all. (493-494)

How *do* we experience shape in touch? Richardson writes: "We are not, in touch, aware of a region of space within which we tactually perceive objects in just this way. We are aware of the things we perceive, tactually, as in contact with the boundaries of an object, namely, one's body" (493). Visual shape experience presents objects as occupying regions within a spatial field of which we are also aware, while haptic experience does not afford awareness of a spatial field within which objects appear to be located.[iv]

Though compelling, I doubt that this view pinpoints the core distinction between visual and haptic shape experience. For it is doubtful that visually experiencing an object's shape necessitates experiencing the space containing the object. Schwenkler (2012) argues that the phenomenon of Bálint's syndrome threatens this view. Bálint's patients can reliably identify certain intrinsic shapes (e.g., letter shapes), but are at chance in reporting the relative positions of two objects (Friedman-Hill et al., 1995). They also exhibit *simultanagnosia*, or the awareness of just one object at a time, with no reported experience of the wider space encompassing the object. This suggests that visual experiences of shape may sometimes occur without experiences of a wider spatial field. If so, then the latter cannot underwrite the core distinction between visual and haptic shape experience.[v]

These comments are not intended as decisive, but I suggest that they are sufficiently worrisome to warrant the search for a different solution to the cross-modal puzzle.

3. Tier 1 Shape Phenomenology

I turn to my positive proposal. I begin with *Tier 1 shape phenomenology*. When theorists speak of the way an object's shape *appears from a perspective*, I think they normally have this tier in mind. Tier 1 phenomenology has three distinctive features.

First, it underdetermines the external, intrinsic shape properties we are aware of in perceptual experience, and vice versa. Two experiences can share Tier 1 phenomenology while differing in the external, intrinsic shapes they represent, and experiences can represent the same external, intrinsic shape while differing in Tier 1 phenomenology.

Second, Tier 1 phenomenology in a modality exhibits similarities that track similarities among the proximal features the modality exploits to recover external, intrinsic shape. Experiences alike in Tier 1 phenomenology tend to be produced by similar proximal stimuli.

Third, our awareness of external, intrinsic shape is *imaginatively entangled* with Tier 1 phenomenology. Although Tier 1 phenomenology does not determine our awareness of intrinsic shape, the two are intimately bound together in the sense that whenever we perceptually imagine something as having some intrinsic shape, our imaginative state exhibits some Tier 1 phenomenology or other.

This section argues that Tier 1 shape phenomenology arises in both vision and touch, but involves awareness of distinct properties in the two cases. For now, I'll only discuss the first two features of Tier 1 phenomenology. I discuss the third in section 6.

3.1. Vision

Consider the wooden circle shown in figures 1a and 1b. We arguably experience the object as circular from both perspectives, but there is also a sense in which its appearance changes. Our experience of the circle thus exhibits *perspectival variation*. More generally, objects perceived as having the *same* intrinsic shape can appear *different* when perceived from different perspectives. We can distinguish perspectival variation from what Lande (2018) calls *perspectival similarity*. When you view the slanted circle, it appears similar in some respect to an ellipse viewed head on (figure 1c), but at least in this respect does not appear similar to the head-on circle. More generally, objects perceived as having *different* intrinsic shapes can appear the *same*, in some respect, from certain perspectives.

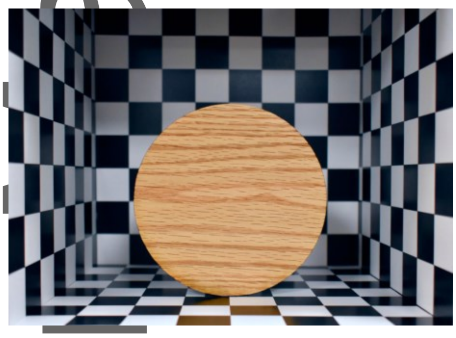

Figure 1a

Figure 1b

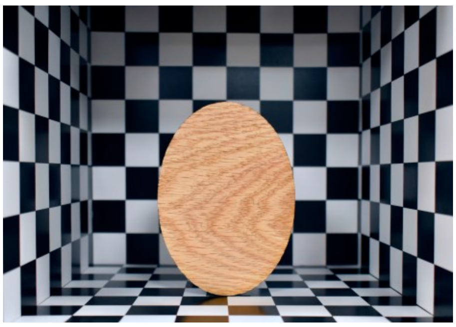

Figure 1c. *Source:* Morales et al. (2020)

Some endorse perspectival variation but not perspectival similarity. Perhaps the head-on circle and the slanted circle appear different, but the slanted circle does not appear similar in any relevant respect to the head-on ellipse (Smith, 2002; Schwitzgebel, 2006, 2011; Hopp, 2013). Rather, visual experience simply represents the head-on circle and slanted circle as differing in orientation relative to the perceiver. This account does not predict any similarity between experiences of the slanted circle and the head-on ellipse. Others endorse both perspectival variation and perspectival similarity (Noë, 2004; Cohen, 2010; Lande, 2018).

If I had to wager, I'd say my experiences of the slanted circle and the head-on ellipse are similar in some respect. However, this claim is controversial (Schwitzgebel, 2006, 590). Moreover, even if we can appreciate some similarity between slanted coins and ellipses, one could question whether this is a genuinely perceptual similarity, or one appreciated only through a special act of imagination (Briscoe, 2008).

A recent visual search study bears on the issue of perspectival similarity. It is known that visual similarity strongly influences visual search: targets take longer to find when they are visually more similar to distractors (Duncan & Humphreys, 1989). Accordingly, Morales et al. (2020) showed participants a circle and an ellipse and simply asked them to judge as rapidly as possible which was the ellipse. The circle was presented either slanted or head-on. Crucially, if the slanted circle appears similar to the ellipse, then, given that visual search is influenced by visual similarity, the ellipse should take longer to identify in this condition. This was so. The effect generalized to real objects shown under full-cue conditions, and also to other shape classes.[vi] Such results suggest that perspectival similarity is real and efficacious in visual search.

Someone might object that perhaps the similarity affecting visual search for ellipses alongside slanted circles was not *phenomenally manifest*, instead residing in low-level visual processing. However, this maneuver seems ad hoc. Other similarities that influence visual search (e.g., in size, color, or shape (Duncan & Humphreys, 1989)) are clearly present in phenomenology, so the objector must explain why this case is an exception. Moreover, studies show that visual search is governed by features recovered at sophisticated levels of perceptual analysis, such as amodally completed contours (Rauschenberger & Yantis, 2001), indicating that search computations are not generally confined to information in low-level processing (Morales et al., 2020, 9).[vii]

So there is introspective and emerging empirical evidence that visual experiences of the slanted circle and head-on ellipse are similar in some respect. I'll assume henceforth that perspectival similarity is real. Tier 1 visual phenomenology comprises those aspects of visual shape phenomenology that exhibit *both* perspectival variation *and* perspectival similarity. Such phenomenology differs between experiences of the head-on circle and the slanted circle, but is shared between experiences of the slanted circle and the head-on ellipse.

How should we account for Tier 1 visual phenomenology? Some views analyze it in terms of our awareness of relational properties of objects—viz., properties objects have as a function of their relation to the perceiver's viewpoint or the perceptual context (Schellenberg, 2008; Hill, 2016; Green & Schellenberg, 2018). Perhaps visual experience represents the *solid angle* an object subtends, which is formed by the set of rays emanating from an apex at the perceiver's viewpoint and just grazing the object's visible boundary (Huemer, 2001, 121; cp. Jagnow, 2012). The slanted circle and head-on ellipse subtend the same solid angle, while the slanted circle and head-on circle do not.[viii]

Others analyze Tier 1 visual phenomenology in terms of the varying *ways* that vision represents intrinsic shape. When we see a slanted coin, our visual state represents circularity under a different mode of presentation (Burge, 2010, 2014), or via a differently structured vehicle (Lande, 2018), than when we see the coin head-on. On Lande's (2018) view, perspectival character is grounded in the array format of visual representations. Visual shape representations are composed of cells, each of which represents features in a given direction; adjacent cells are dedicated to adjacent directions. Because the boundaries of the head-on ellipse and the slanted coin fall along the same directions, they are assigned the same cells. Conversely, the slanted circle and head-on circle are assigned different cells.

I won't dwell on the differences between these approaches. Rather, I highlight two common elements. First, both approaches hold that Tier 1 visual phenomenology underdetermines the intrinsic shapes represented in experience, and vice versa. Objects experienced as having the same intrinsic shape can be experienced as subtending different solid angles, or can be assigned different cells in one's representational array. Likewise, objects experienced as differing in intrinsic shape can be experienced as subtending the same solid angle, or can be assigned the same cells in one's representational array.

Second, both approaches predict that Tier 1 visual phenomenology exhibits similarities that track similarities in the proximal stimulus. The retinal image compresses with shifts in perspective, and changes in solid angle mirror these compressions. Likewise, if visual arrays comprise cells dedicated to particular lines of sight, then objects with the same retinal image are assigned the same cells of the array. If Tier 1 visual phenomenology is grounded in awareness of either of these sorts of properties, then we should expect it to exhibit similarities and differences that mirror the proximal stimulus for vision.[ix]

For convenience, I'll assume that Tier 1 visual phenomenology involves awareness of solid angle. This is just a placeholder for whatever feature underwrites the apparent similarity between slanted circles and head-on ellipses—be it a property represented in experience or an internal structural feature of experience. My arguments will be neutral on this issue. However, the distinction is important for other reasons. It specifically bears on whether representationalism (see section 6) can accommodate Tier 1 visual phenomenology. If Tier 1 phenomenology is grounded in relational properties represented in experience, then it can. But if it is grounded in properties *instantiated* by experiences, then it arguably cannot. I set this dispute aside for present purposes.

3.2. Touch

The proximal stimulus for touch is not produced through optical projection, so it avoids the kinds of compression found in the retinal image. Suppose you first pick up a circular coin oriented head on, and then you pick up a same-shaped coin at an angle. The coins apply pressure patterns that stimulate mechanoreceptors under the skin. These pressure patterns form the proximal stimulus used by the haptic system to recover the coin's shape. However, the distances between pressure points are the same in both cases because your hand reorients to grasp the coin depending on its slant. The haptic proximal stimulus does not systematically compress with slant.

Nonetheless, haptic shape experience generates analogous puzzles of perspectival variation and perspectival similarity. Suppose you explore a coffee mug with your right hand, and then you pass it to your left hand. Later, you explore the mug with your foot. These shifts affect the mug's haptic appearance, but it also appears to remain the same intrinsic shape and size. This generates a puzzle of perspectival variation. You experience the mug as the same intrinsic shape regardless of whether you hold it in your right or left hand, but it also appears different between these perspectives (see also Schwenkler, 2019, 275-276; Matthen, forthcoming). So haptic shape experience exhibits perspectival variation.

Figure 2a

Figure 2b

It also exhibits perspectival similarity. Suppose you hold the objects in figures 2a and 2b in succession. In both cases, you grasp the large square part with your left hand, and you grasp the smaller elliptical part with your right hand. You use the same parts of your hands to explore the object in both cases, but the relative orientation of your hands changes. The same body parts are in contact with the objects, but their apparent intrinsic shapes differ because the haptic system also exploits proprioceptive information about body posture and movement in computing shape (Yau et al., 2016; Briscoe, 2019). There is a respect in which the two objects haptically appear similar from one's perspective, but you experience them as differing in intrinsic shape.

Tier 1 haptic phenomenology requires awareness of one's body. We must, however, distinguish it from tactile sensations like itches or aches. Tactile sensations may represent the condition of certain body parts (Tye, 2000, 50), but they don't represent the presence of an *external object* impinging on those body parts (see Matthen, forthcoming). Tier 1 haptic phenomenology is *externally directed* in a way that mere tactile sensations are not.

I propose that Tier 1 haptic phenomenology involves experientially representing an external object as applying pressure to a collection of body parts $P_1 \ldots P_n$. Such pressure can result from the voluntary use of force, but need not. We can experience an object as applying pressure to certain body parts while we remain passive and motionless, and we can perceive the geometrical features of objects under these conditions (Haggard & Giovagnoli, 2011).

The applied pressure can be either *direct* or *indirect*: It can involve immediate bodily contact with an object (as is typical), or it can involve a material intermediary between the object and one's body. The latter provision accommodates cases of what Fulkerson (2014) calls *distal touch*. Suppose that you become aware of the shape of a rock by exploring it with a stick. Here, the object is not in direct contact with your body, but applies pressure to your body by way of the stick.[x]

We can understand the collection of body parts to which an object applies pressure as the object's *somatotopic projection*—the locations, in somatotopic coordinates, where the object presses against the body. Somatotopic coordinates remain fixed despite shifts in body posture. As I wave an object back and forth, its somatotopic projection remains stable. A *spatiotopic* reference frame, conversely, specifies the locations where an object pressures the body in external coordinates that shift with changes in posture (cp. Bermudez, 2005). The haptic system employs both somatotopic and spatiotopic encoding. Subjects are worse at judging the order in which their hands are tapped when their arms are crossed (Yamamoto & Kitazawa, 2001), likely due to the conflict between left-right relations in somatotopic and spatiotopic frames. Stimuli are encoded somatotopically during the first 100 ms after stimulus onset, while spatiotopic encoding emerges 200 ms post-onset and recruits posterior parietal cortex (Azañón & Soto-Faraco, 2008; Azañón et al., 2010). Tier 1 haptic phenomenology plausibly arises from earlier stages of somatotopic encoding.

It is possible that the experiential representation of somatotopic projection accounts for only *part* of Tier 1 haptic phenomenology. Perhaps Tier 1 haptic phenomenology *also* involves the experiential representation of bodily posture or movement.[xi] For example, your haptic perspective on the shape of a pencil might differ depending on the specific exploratory movements you use to recover its shape, or the specific hand posture you adopt. I find these suggestions plausible, and I am inclined to believe that Tier 1 haptic phenomenology is internally complex. Thus, my central claim is that the experiential representation of somatotopic projection constitutes a *core element* of Tier 1 haptic phenomenology, and that it determines the specific patterns of perspectival variation and similarity discussed earlier. Other aspects of Tier 1 haptic phenomenology might generate their own, independent patterns of perspectival variation/similarity.

Before continuing, let me forestall a concern with the idea that shifts in somatotopic projection suffice for shifts in haptic perspective. It's clear that changes in an object's somatotopic projection alter the proximal stimulus for touch. However, not all changes in the proximal stimulus produce changes in perspective on an object's shape. Suppose you view a coin head-on with your left eye closed, then view it again with your right eye closed. This changes the proximal stimulus for vision, but doesn't seem to alter your perspective on its shape in any meaningful way. So why should we construe changes in somatotopic projection as analogous to changes in the coin's orientation as opposed to changes in the eye used to view it?

My reply appeals to a disanalogy between rotating the coin and viewing it with different eyes. Although both alter the proximal stimulus for perception, arguably only the first alters the *inputs* to visual computations of shape. The inputs to visual shape computations certainly include an object's 2D retinal shape (e.g., Pizlo, 2008), though other cues are likely important too.[xii] Critically, 2D retinal shape remains the same through mere changes in which retina an object stimulates (assuming the object is oriented the same way vis-à-vis both eyes), but shifts with changes in orientation. So information about which eye is used to view the coin may be filtered out prior to the computation of external shape. In the haptic case, however, information about somatotopic projection *cannot* be filtered out this way. Haptic computations of shape need to integrate information about pressure to the body with proprioceptive information about body posture. Because the same proprioceptive information signifies different external shapes depending on how an object pressures the body (Berryman et al., 2006; Briscoe, 2019; Matthen, forthcoming), somatotopic projection is *ineliminable* in the haptic computation of shape. Perhaps, then, the proximal changes that mark changes in perspective on a property are just those changes that affect the inputs to computations responsible for recovering the property. If so, then changes in somatotopic projection are more like changes in the coin's orientation than in the eye used to view it.

What's the relation between Tier 1 visual phenomenology and Tier 1 haptic phenomenology? I argue that the two are analogous but distinct.

Our awareness of somatotopic projection underdetermines the intrinsic shapes we perceptually represent, and vice versa. Tier 1 haptic phenomenology is shared between experiences of differently shaped objects explored using the same body parts, and varies between experiences of same-shaped objects explored using different body parts. This mirrors the way visual awareness of solid angle underdetermines visual awareness of intrinsic shape, and vice versa. So Tier 1 phenomenology is analogous in this respect between sight and touch.

Furthermore, Tier 1 haptic phenomenology exhibits similarities that track similarities in the proximal features used to recover an object's intrinsic shape. The pressure forces applied to the body simply *are* the proximal stimulus for haptic shape recovery. So if Tier 1 haptic phenomenology involves awareness of an object's somatotopic projection, then trivially it mirrors the proximal stimulus for touch. In this respect, too, Tier 1 phenomenology is analogous across modalities.

Nonetheless, Tier 1 shape phenomenology differs across modalities. We only experience an object's solid angle through vision, and we only experience an object's somatotopic projection through touch. Because these properties are different, they determine distinct *patterns* of perspectival similarity and variation between modalities. Two objects can subtend the same solid angle while differing in their somatotopic projection, and vice versa. Thus, Tier 1 shape phenomenology is analogous but distinct across modalities.

4. Tier 2 Shape Phenomenology

Tier 2 shape phenomenology is perspectival because it changes with shifts in perspective on an object. However, unlike Tier 1, Tier 2 phenomenology is not shared between experiences of objects that we perceive as differing in intrinsic shape. Tier 2 phenomenology thus exhibits perspectival variation but not perspectival similarity. I'll also argue that Tier 2 phenomenology differs across modalities.

Tier 2 phenomenology involves awareness of an object's intrinsic shape within a perceiver-centered reference frame, where parts of an object are specified by their distances and orientations relative to an origin and axes anchored to the perceiver's body. The idea that visual experience represents shape in a perceiver-centered manner is not new (Peacocke, 1992, 63; Briscoe, 2008; Bennett, 2012). My present claim goes beyond this. I suggest that perceiver-centered shape representations arise in touch as well as vision, but that the associated phenomenology differs in the two cases due to differences in the relevant perceiver-centered frames.

We should distinguish the question of whether visual experience represents intrinsic shape *at all* from the question of whether the experiential representation of intrinsic shape is *invariant to perspective*. An object's intrinsic shape consists in the geometrical configuration of its boundary points. However, the same geometrical configuration can be represented in different reference frames, and within some frames, the way it is represented varies with the object's position or orientation. A perceiver-centered representation as of a slanted circle, for instance, differs from a perceiver-centered representation as of a head-on circle because the representations attribute different distances, directions, and orientations to parts of the circle. Nonetheless, both representations can fully specify the circle's intrinsic shape. Perceiver-centered shape representations represent intrinsic shape, but the *way* they represent intrinsic shape varies with shifts in perspective. Hence they exhibit perspectival variation.

Recall that two experiences can represent the same solid angle or somatotopic projection while differing in the intrinsic shapes they represent. Conversely, any two experiences that represent the same perceiver-centered shape content (the same arrangement of distances, directions, and orientations of an object's edges and surfaces) also represent the same intrinsic surface shape. Thus, Tier 2 phenomenology does *not* exhibit perspectival similarity. It is not shared between experiences of objects we perceive as differing in intrinsic shape.

4.1. Vision

The images in figure 3 all subtend the same solid angle, but they clearly appear different.[xiii] Plausibly, *one* way they differ is that the objects' surfaces appear to be arranged differently in depth relative to the perceiver. We experience certain regions as protruding toward us, others as receding away, and these patterns vary among the objects. At some level, visual experience seems to represent an object's shape in terms of the distances and orientations of its component surfaces and edges. This is Tier 2 visual phenomenology.

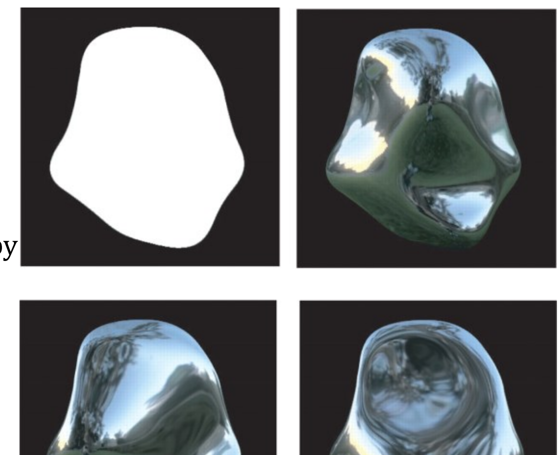

Figure 3. *Source:* Fleming et al. (2004). Reproduced with kind permission of the Association for Research in Vision and Ophthalmology.

One might object to this introspective argument. To be sure, we visually experience the perceiver-centered distances and orientations of *local surface patches*. However, one might insist that this is distinct from the phenomenology of *shape*—which concerns an object's global geometrical structure. Perhaps the visual experience of shape comprises both perspectival Tier 1 aspects and *object*-centered shape content (see section 5), but no intermediate layer of perceiver-centered content.

However, there is compelling evidence for visual representations of shape in perceiver-centered coordinates. Such representations are involved in recognizing objects through changes in orientation. To play this role, the representations must integrate information about surface patches into a specification of the object's overall shape. Since these representations are employed in the prototypically ventral process of object recognition, it is reasonable to suppose that they ground aspects of visual shape phenomenology (Goodale & Milner, 1992).

Visual recognition is view-dependent: When we see a shape at one orientation, we are often worse at recognizing the shape at new orientations than at the familiar one (Tarr & Pinker, 1989; Peissig & Tarr, 2007). Several studies suggest that these patterns of view-dependence are best explained by representations of global shape in perceiver-centered coordinates.[xiv] Bennett and Vuong (2006) found that when subjects were familiarized with an object at one orientation, they recognized it more accurately at new orientations when stereo cues were available than when they were absent. However, there was still a significant *drop* in performance at new orientations even in the stereo condition (cp. Cristino et al., 2015). To explain the stereo advantage, the shape representations employed must encode information about distance, not merely 2D image features (*pace* Poggio & Edelman, 1990). However, they must also encode shape differently at different orientations to explain the falloff in performance with orientation changes. Representations of intrinsic shape via perceiver-centered distance and orientation can explain both findings.

So the earlier introspective observation converges with psychophysical evidence. The visual system forms representations that specify intrinsic shape in terms of perceiver-centered features. These representations underlie Tier 2 visual phenomenology. Such phenomenology exhibits perspectival variation but not perspectival similarity. The slanted circle and head-on circle differ at Tier 2 because their edges and surfaces are experienced as having different depths and orientations. However, the slanted circle *also* differs from the head-on ellipse in these respects, so perspectival similarity is absent. Thus Tier 2 phenomenology contrasts with Tier 1.

4.2. Touch

Sensitivity to viewpoint might seem like a distinctively visual phenomenon. When we see an opaque object from a fixed viewpoint, its backsides are hidden from view, making it more difficult to recover its entire shape. Conversely, when we feel an object we are often free to explore most of its surfaces at a single orientation. Occlusion is a more pervasive issue in vision than in touch. Interestingly, however, there is evidence that perceiver-centered shape representations are formed and used in haptic recognition as well. I conjecture that such representations underlie aspects of our haptic experience of shape.

I start with evidence for orientation-dependence in haptic recognition. Lacey et al. (2007) had participants feel a set of four objects each composed of a collection of blocks. Each object was explored at a fixed orientation, although subjects were free to explore both its back and front. Next, a single object was presented and subjects had to say which of the initial four it was. When the object was presented in a new orientation, recognition accuracy dropped significantly, mirroring standard results from vision. Orientation-dependent haptic recognition has been documented in several other studies as well (Newell et al., 2001; Lacey et al., 2009; Lawson, 2009, 2011). Thus, the haptic system plausibly forms representations of shape in perceiver-centered reference frames. Such representations exhibit perspectival variation and account for recognition costs across orientation changes.

So there is a level at which haptic shape representation is perceiver-centered. But what sort of perceiver-centered reference frame is involved? Where is its origin, and what are its axes?

Astrid Kappers and colleagues have shown that systematic biases in the perception of geometrical relations during manual exploration are best explained within a *hand-centered* frame (Kappers & Viergever, 2006; Volcic & Kappers, 2008; van Mier, 2014), where lines perceived as "vertical" are, roughly, those parallel to the imaginary line leading from the center of the wrist to the tip of the middle finger. Kappers and Viergever (2006) blindfolded participants and asked them to adjust the orientation of a test bar until it felt parallel to a reference bar felt with the other hand. Subjects had to keep the relative orientations of their hands fixed during the task: both hands pointed outward, both pointed rightward/leftward, converging toward each other, or diverging away. Kappers and Viergever found that errors varied predictably with hand posture. When subjects' hands diverged and they felt a vertical reference bar with their right hand, the apparently parallel test bar was rotated about thirty degrees counterclockwise from the reference bar. While the bars were objectively non-parallel, they were oriented approximately the same way relative to their respective hands.[xv] If subjects tried to make the bars parallel by positioning them at roughly the same hand-centered orientation, one would expect this sort of error. Other perceiver-centered frames (e.g., eye-centered, trunk-centered) cannot explain this result.[xvi]

These findings highlight the difference between Tier 1 and Tier 2 haptic phenomenology. When the reference/test bars were held with different hands, their Tier 1 phenomenology differed. However, they were also experienced as *sharing* perceiver-centered features—namely, the property of being a straight rod oriented thus-and-so. The latter is a case of Tier 2 haptic shape phenomenology. Another example: Place your cell phone in front of you so that its vertical axis runs outward. Now close your eyes and explore it with your right hand, and then with your left. Next, turn the phone sideways and explore it with your right hand again. In an important respect, the first two experiences are more similar than either is to the third. The distinctive phenomenology of feeling your phone's shape at a vertical hand-relative orientation arises regardless of which hand you use to feel it.

Visual and haptic Tier 2 phenomenology differ in reference frame. We are visually aware of an object's structure relative to our eyes and direction of gaze—its layout in "visual space"—and haptically aware of its structure relative to the body part(s) used to explore it—its layout in "haptic space." Thus, Tier 2 phenomenology typically differs across modalities.

Must it *always* differ? Might you orient your body so that an object's layout in haptic space precisely matches its layout in visual space? If so, wouldn't the same perceiver-centered shape properties be represented both visually and haptically? If so, this would suggest that Tier 2 phenomenology *can* match cross-modally, at least on occasion. I'll remain agnostic on whether this can occur. For it to occur, however, visual and haptic egocentric reference frames must be the same *except* for the placement of their origins and axes. They needn't be. If, say, one modality employs polar coordinates while the other employs Cartesian coordinates, there would be no way to position the object so that its perceiver-centered representations precisely matched cross-modally. In any case, we can conclude that Tier 2 phenomenology *usually* differs across modalities. It may always differ, but even if not, special effort would be required to produce a match.

5. Tier 3 Shape Phenomenology

I've elucidated two tiers of shape phenomenology that are both perspectival and modality-specific. This section turns to *Tier 3 shape phenomenology*. Unlike its predecessors, Tier 3 exhibits neither perspectival similarity nor perspectival variation. Rather, it involves awareness of an object's intrinsic shape in object-centered coordinates that remain fixed through perspective shifts. I'll also argue that it is constitutively multisensory.

5.1. Tier 3 Phenomenology in Vision

Structural schemes of shape representation have played a prominent role in perceptual psychology. This subsection argues that the contents of structural representations are reflected in visual shape experience. Such representations underpin Tier 3 shape phenomenology.

Two characteristics of structural representations deserve emphasis.[xvii] First, they are *part-based*. A coffee mug is represented as comprising a cylindrical part and a handle; a human body is represented as comprising head, limbs, and torso. Structural schemes privilege these parsings over others because part decomposition follows systematic principles. The *minima* rule says that part boundaries are located at sharp curvature concavities, while the *short-cut* rule says that an object is divided into parts by linking these concavities along the shortest paths possible (Hoffman & Richards, 1984; Singh & Hoffman, 2001). Part structure thus defined is efficacious in perceptual processing: It influences visual attention (Barenholtz & Feldman, 2003), visual search (Xu & Singh, 2002), the perception of transparency (Singh & Hoffman, 1998), and the perception of orientation (Cohen & Singh, 2006).

The second characteristic is *view-invariance*. The intrinsic shapes of an object's parts are encoded relative to axes anchored to the parts, and the spatial relations between parts are encoded in terms of the parts' relations to one another, rather than to the perceiver's viewpoint (Biederman, 1987; Hummel, 2013). For example, one part might be represented as *atop* or *side-attached* to another.

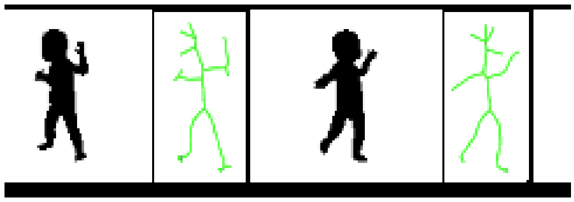

Figure 4a. Medial axis structure of the hand. *Source:* Feldman & Singh (2006). Copyright 2006 National Academy of Sciences.

Figure 4b. Medial axis structure of the body. *Source:* Kimia (2003). Reproduced with kind permission from Elsevier.

A growing body of evidence suggests that the visual system constructs view-invariant representations by extracting an object's *medial axis structure*. The medial axis is composed of the set of points in the interior of a figure that have two or more closest points on the figure's boundary. The medial axis resembles a "skeleton" from which the object is grown. Medial axes figure prominently in structural schemes because distinct parts of an object normally correspond to distinct axis branches (Kimia, 2003). The medial axis of a human hand includes separate branches for each finger, and the medial axis of the body has separate branches for separate limbs (figs. 4a-4b). These axes can be incorporated into an object-centered representation wherein edges along a shape's boundary are encoded via their distances and directions from corresponding points along its axis (Blum & Nagel, 1978; Feldman & Singh, 2006). The relations between parts are specified through the locations at which different axis branches intersect one another. This representation remains fixed through changes in position and orientation (i.e., is view-invariant).

Our visual systems recover medial axis structure and exploit it in sensory processing. Medial axis structure influences visual contrast sensitivity (Kovács et al., 1998) and texture segregation (Harrison & Feldman, 2009). Furthermore, when subjects are asked to tap a shape wherever they like, taps tend to cluster around the medial axis in ways that cannot be explained by a general tendency to tap shapes toward the center (Firestone & Scholl, 2014; Ayzenberg et al., 2019). Finally, physiological evidence indicates that the lateral occipital cortex (LOC), an area dedicated to representing shape for object recognition, encodes medial axis structure in an orientation-invariant manner. Patterns of activity are selective for an object's medial axis structure, and these patterns remain relatively stable through changes in the object's orientation (Hung et al., 2012; Lescroart & Biederman, 2013).

Such evidence suggests that structural shape representations are probably generated in visual processing. Are their contents reflected in visual phenomenology? I believe they are.

Structural representations can explain the stability in shape appearance through changes in orientation. Consider the coin. The coin's medial axis includes just a single point (its center), and its outer boundary is specified by its radius. On the current proposal, your experience exhibits stable Tier 3 shape phenomenology despite changes in the coin's orientation because you experience it as possessing a fixed medial axis structure and fixed radius throughout.

The role of structural representations is even more salient in experiences of complex, multi-part objects. Because such representations make part structure explicit, they enable a distinction between changes that leave an object's part shapes intact and those that alter part shapes, even when both involve changes in the object's global shape, position, and orientation. When a person walks across the room, her global shape changes constantly, as do her perceiver-centered position and orientation. However, a structural scheme explicitly represents the features that remain stable through such changes: the decomposition of the body into parts, the intrinsic shapes of those parts in part-centered coordinates, and the part-centered locations of joints between parts. These features together comprise an object's *compositional structure* (Green, 2019).

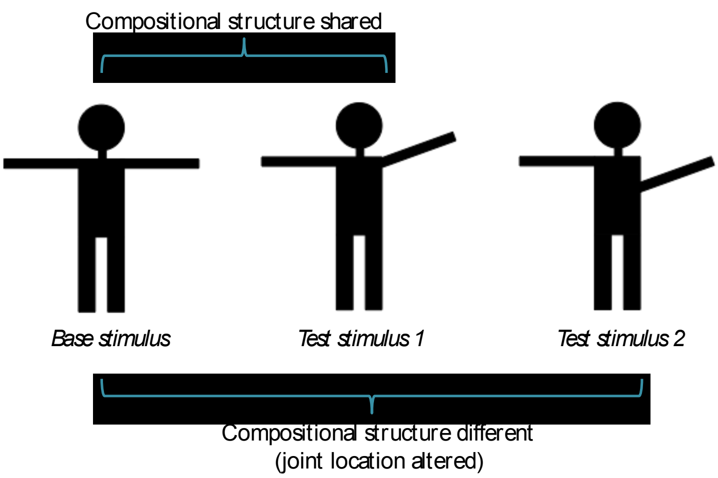

Figure 5

All else equal, changes that disrupt an object's compositional structure are more salient or striking than those that preserve it, and this difference is phenomenally manifest. Consider the two transformations of the human body shown in figure 5. Test stimulus 1 shares compositional structure with the base stimulus on the left, while test stimulus 2 does not, since a joint location is altered. The change from the base to test stimulus 2 is more phenomenally salient than that from the base to test stimulus 1. Here apparent shape similarity is governed by compositional structure.

Psychophysical results substantiate this introspective observation and extend it to novel objects where familiarity is not a factor. Barenholtz and Tarr (2008) showed subjects a novel base shape along with two transformations of the base, one of which retained compositional structure while the other did not (involving either a change in joint location or in intrinsic part shape). When asked which of the two transformed shapes was more similar to the base, subjects reliably selected the shape that preserved compositional structure, even though both had roughly the same pixel-by-pixel overlap with the base shape.

More recently, Lowet et al. (2018) found similar results in a shape discrimination paradigm. Subjects saw a pair of shapes and judged whether they were the same or different. On same-shape trials, one object was a rotated version of the other, while on different-shape trials, the objects either shared or differed in medial axis structure. Shapes that shared medial axis structure tended to share part shapes and part-centered joint locations, but differed in the angles between parts. Lowet et al. found that subjects were significantly better at detecting shape differences across rotation when the objects differed in medial axis. Thus, shapes that share compositional structure are less perceptually discriminable than those that don't; this pattern holds across changes in position and orientation.

So compositional structure influences apparent shape similarity. This, I claim, is because it is represented in visual experience. Objects that share compositional structure are experienced as sharing a property; those that differ in compositional structure are not.

Tier 3 phenomenology involves awareness of part-based, view-invariant structure. Thus, it exhibits neither perspectival similarity nor perspectival variation. It involves awareness of intrinsic shape, so it is not shared between experiences of the slanted coin and head-on ellipse (unlike Tier 1). And it involves awareness of shape within view-invariant, object-centered reference frames, so it remains stable through shifts in perspective (unlike Tiers 1 and 2).

5.2. Tier 3 Phenomenology is Constitutively Multisensory

Introspection seems to reveal Tier 3 phenomenology in touch as well as vision. When you explore a mug with your eyes closed, the handle seems like a separate part from the cylinder, and you experience its global shape as remaining stable despite changes in its orientation or in the hand you use to explore it. We also seem to form expectations based on haptic experiences of part structure. When you feel a toy action figure, it seems more likely that the figure's limbs will turn about their joints than that they will stretch or contract. But are these intuitions empirically supported? If so, is Tier 3 shape phenomenology the same or different across modalities?

I claim that Tier 3 shape phenomenology is *constitutively multisensory*. It is not uniquely associated with either vision or touch. Rather, at least sometimes it belongs to both modalities. But what is it for an aspect of phenomenal character to be uniquely associated with a modality? O'Callaghan (2015, 2019) helpfully distinguishes two answers.

First, we might say that an aspect of phenomenal character—a "phenomenal feature"—is uniquely associated with a modality when it is *distinctive* to that modality. Phenomenal feature P is uniquely associated with modality M when no experience that is not of M could instantiate P. If so, the claim that Tier 3 shape phenomenology is constitutively multisensory amounts to the claim that the features characterizing Tier 3 phenomenology can be instantiated by both visual and haptic experiences.

A second approach allows that the same phenomenal feature could be instantiated by a visual experience at one time and by a haptic experience at another time, but be uniquely associated with just one modality on each occasion. On this approach, the phenomenal features associated with a modality on an occasion include "just that which a *corresponding mere perceptual experience* of that modality could instantiate" (O'Callaghan, 2015, 561, emphasis added). A *mere* perceptual experience of modality M is an experience of M that is not an experience of any other modality—e.g., a visual experience that isn't auditory, haptic, gustatory, or olfactory. A *corresponding* mere experience of modality M is a mere experience of M "under equivalent stimulation" (2019, 129). Suppose that a subject enjoys concurrent visual and haptic phenomenology thanks to retinal stimulation R and cutaneous stimulation C. Then, O'Callaghan suggests, a corresponding merely visual experience is one that the subject could have undergone had she received R alone on that occasion. On this conception, Tier 3 shape phenomenology is constitutively multisensory provided that some perceptual experiences have Tier 3 phenomenology that no corresponding merely visual or merely haptic experience could have.

Call the first conception the *distinctiveness conception*, and the second the *unisensory-correspondence conception*. They differ. It's possible that Tier 3 phenomenal features are instantiated by both visual and haptic experiences at various times, but that whenever they occur, they could have been instantiated by a corresponding merely visual or merely haptic experience. Then Tier 3 phenomenology would be multisensory under the distinctiveness conception but not the unisensory-correspondence conception.

I argue that Tier 3 shape phenomenology is constitutively multisensory under both conceptions. The same Tier 3 phenomenal features can be instantiated by both visual experiences and haptic experiences, so Tier 3 phenomenology is distinctive to neither modality. Furthermore, some experiences exhibit Tier 3 phenomenal features that no corresponding merely visual or merely haptic experience could exhibit.

5.2.1. The Distinctiveness Conception. Leading views of the metaphysics of perceptual experience divide into two camps. Some accept what Pautz (2017) calls *internal dependence*. These views hold that the phenomenal character of experience is determined with at least nomic necessity by physical-functional properties of the brain. Such views include type physicalism (Hill, 1991), narrow functionalism, Pautz's brand of non-naturalistic representationalism (Pautz, 2017), and forms of property dualism on which mental properties nomically supervene on internal physical-functional properties (Chalmers, 1996). Others reject internal dependence, holding instead that phenomenal character is grounded in the worldly properties of which we are aware. Naïve realists hold that we become aware of such properties by being directly acquainted with them (Campbell, 2002; Brewer, 2011), while naturalistic representationalists hold that we become aware of them when they figure in the representational content of experience (Dretske, 1995; Byrne & Tye, 2006). I argue that regardless of which camp you occupy, you should accept that the same Tier 3 phenomenology can be instantiated by both visual and haptic experiences.

Suppose, first, that you believe phenomenal character is grounded in the worldly properties we perceive. To motivate the claim that Tier 3 phenomenology is multisensory on this view, we must show that both modalities afford awareness of the same part-based, view-invariant structure. Evidence suggests that this is true.

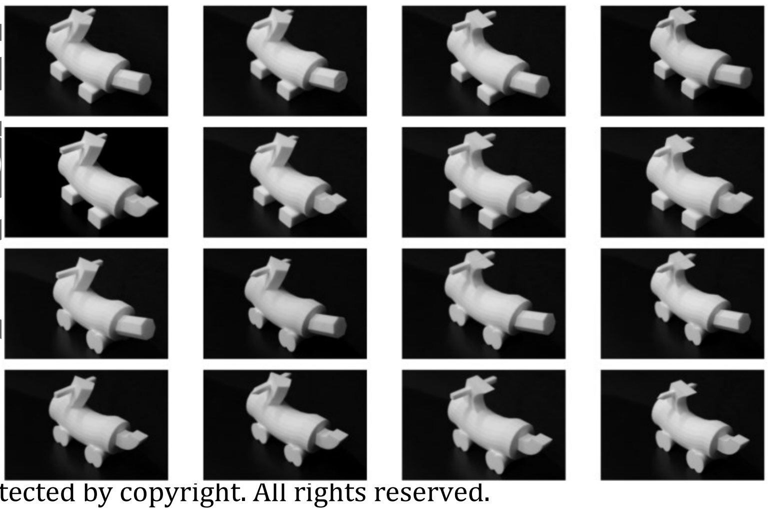

Figure 6. *Source*: Erdogan et al. (2015). Reproduced under the Creative Commons Attribution License.

First, both modalities appear to represent shape in a part-based manner, and they parse objects in similar ways. Erdogan et al. (2015) produced a set of 16 objects (fig. 6). All objects had the same overall spatial arrangement, but individual parts could adopt one of two shapes. Subjects rated the similarity of each pair of objects from 1-7. On a given trial, the objects were either both seen, both felt, or one seen and the other felt. Similarity ratings were strongly predicted by shared part shapes, and average ratings were highly correlated (~0.95) both between modalities and between intramodal and cross-modal conditions (see also Gaißert et al., 2010; Masson et al., 2016). Plausibly, then, we apprehend the same part structure in vision and touch, and this helps explain why shape similarity ratings are similar regardless of modality.

Second, studies on cross-modal recognition suggest that vision and touch both represent shape in a view-invariant manner. The Lacey et al. (2007) study discussed earlier found that recognition is view-dependent within both vision and touch. However, the study also investigated cross-modal recognition. Participants encountered four objects in one modality, then had to identify one of the four objects through the other modality, either at the same or a different orientation. Critically, cross-modal recognition was orientation-*invariant*—i.e., equally accurate at familiar and unfamiliar orientations. Subsequent studies have reproduced this result (Lawson, 2009; Lacey et al., 2010; Ueda & Saiki, 2012). Thus, cross-modal recognition seems to employ representations that permit equivalent recognition performance regardless of orientation. A natural explanation is that, at some stage, vision and touch both represent shape in object- or part-centered coordinates, and cross-modal recognition relies on awareness of this same view-invariant structure in both modalities.

These studies suggest that the representations responsible for our ability to *compare* shapes across modalities are structural: part-based and view-invariant. This is why cross-modal similarity ratings are predicted by part structure, and why the ability to recognize shapes across modalities is insensitive to orientation.

Thus, vision and touch plausibly afford awareness of the same part-based, view-invariant structure. If you believe that Tier 3 phenomenal character is fixed by the worldly properties we are aware of, then you should accept that it can be instantiated by both visual and haptic experiences. If so, Tier 3 phenomenology is distinctive to neither vision nor touch.

Now suppose you endorse internal dependence. You believe that phenomenology is fixed by internal physical-functional properties, not by the worldly properties we are aware of. Such views permit modal differences in Tier 3 shape phenomenology even if we are aware of the same properties in both cases, provided that physical-functional differences are present.

However, the physical-functional states responsible for our awareness of part-based, view-invariant structure probably do not vary between modalities. The very same internal states represent these properties in both cases. Thus, if you accept internal dependence, then you should *also* accept that Tier 3 phenomenology is distinctive to neither vision nor touch.

Section 5.1 argued that Tier 3 visual phenomenology is underpinned by structural representations based partially in the LOC. While LOC was traditionally construed as a purely visual area, more recent studies show that it contributes to haptic shape processing too. LOC is responsive when subjects feel an unseen object with a coherent shape but not when they feel a texture pattern (Amedi et al., 2001). The relevant regions of LOC are not activated by meaningful auditory signals (Amedi et al., 2002), suggesting that it is involved in the multisensory processing of *shape*, not just any information pertinent to recognition. Furthermore, damage to LOC impairs both visual and haptic object recognition (James et al., 2006).[xviii]

Recall that Erdogan et al. (2015) found that part structure governs shape similarity both within and between modalities. A follow-up study investigated the physiological basis of this effect. Erdogan et al. (2016) had subjects either view or feel each of the objects from figure 5 while fMRI recordings were taken from sensory areas. Critically, it was possible to decode the set of part shapes comprising an object from the LOC response it elicited, regardless of whether the object was seen or felt. Furthermore, in both modalities, there was a significant correlation between neural similarity matrices (the degree of similarity between LOC responses to pairs of objects) and the subjective similarity ratings reported by Erdogan et al. (2015). These results suggest that LOC generates part-based representations in both vision and touch, and these plausibly underlie the patterns of apparent shape similarity discussed above.

Lacey and Sathian (2014) propose that view-invariant cross-modal recognition is mediated by LOC. Shape processing in vision and touch begins in unisensory areas, but converges in cross-modal areas later on. Given that LOC is involved in both visual and haptic shape representation, it is a prime candidate for the sort of coordination involved in cross-modal recognition. Moreover, there is evidence that, at least in vision, LOC can produce view-invariant responses to shape (James et al., 2002). And, recall, there is evidence that LOC represents medial axes in a view-invariant manner.

So LOC plausibly supports structural representations in both sight and touch. The final step is to argue that it supports the *same* structural representations in both cases. The mere fact that LOC is activated both visually and haptically does not establish this. The neural states within LOC might differ cross-modally. Even if they are the same, they might play different functional roles in the two cases, thereby implementing distinct representations. I claim both of these possibilities should be rejected.

In the study mentioned earlier, Erdogan et al. (2016) found similar LOC activity across modalities. For all participants, the LOC activity patterns while seeing a shape were significantly more alike when touching the same shape than when touching a different shape. Further evidence for cross-modal continuity in LOC representations comes from neural adaptation. Tal and Amedi (2009) showed subjects pictures of objects, and then had them feel either a same-shaped object or a different-shaped object. fMRI recordings revealed a significantly greater reduction of LOC activity when the same shape was repeated across modalities. This again suggests that LOC exhibits similar response patterns to shape regardless of modality. Because of this, adaptation transfers better across modalities for same-shaped objects than for different-shaped objects.

These findings suggest that the same physiological states are responsible for our awareness of part-based, view-invariant structure in both vision and touch. Still, someone might object that this leaves open the possibility of an internal *functional* difference across modalities. Perhaps the same neural states play different functional roles between modalities, just as pieces of computer hardware implement distinct representations, marked by different functional roles, depending on the program being executed. Against this, I contend that the visual and haptic representations responsible for cross-modal recognition overlap both physiologically and functionally (cp. Green, 2020).

If visual and haptic view-invariant representations share functional role, then revising one representation should mandate the same revisions to the other. Such automatic cross-modal transfer would suggest that the representations are updated and modified in the same way given sensory input. Alongside the physiological evidence already discussed, this would provide strong indication that our visual and haptic view-invariant shape representations are both physiologically and functionally type-identical.

While intra-modal shape recognition generally starts out view-dependent, viewpoint sensitivity decreases over time. There are various reasons this might be so. However, one attractive explanation is that extra familiarization with an object increases the detail, precision, or speed of formation of its view-invariant shape representation, rendering recognition less view-dependent. Such changes are indicative of how view-invariant representations are updated given sensory input. Importantly, a study by Lacey et al. (2009) suggests that whatever revisions are involved in this process transfer automatically across modalities.

In the first stage of the study, subjects encountered novel shapes within a modality and had to recognize them within the same modality. Unsurprisingly, recognition accuracy was orientation-dependent regardless of modality. Next, there was a brief training session within a single modality where subjects either saw or felt the objects at multiple orientations. Finally, recognition was re-tested. Importantly, irrespective of which modality was used in training, recognition became view-invariant within *both* modalities. Training in vision induced view-invariance in touch, and vice versa. Updating view-invariant representations in one modality mandated the same updates to view-invariant representations in the other, suggesting that they don't merely overlap physiologically—they share functional role as well.

Thus, there is evidence that our awareness of part-based, view-invariant structure arises from the very same physical-functional states in both modalities. If you endorse internal dependence, then you should accept that the phenomenology associated with these states of awareness is the same in both cases. If so, then Tier 3 phenomenology is distinctive to neither vision nor touch. On the distinctiveness conception, Tier 3 phenomenology is constitutively multisensory. This holds regardless of whether phenomenal character is fixed by the brain or by the world.[xix]

5.2.2. The Unisensory-Correspondence Conception. I turn to the second conception of what makes a phenomenal feature constitutively multisensory. Under the unisensory-correspondence conception, Tier 3 phenomenology is constitutively multisensory on an occasion provided that it is not shared by any corresponding merely visual or merely haptic experience.

If vision and touch employ the same system of structural representations, it stands to reason that we should be able to combine visual information about one part of an object with haptic information about another part, producing a composite representation of the object's global shape. Suppose you see a coffee mug while its handle is occluded from view. Now reach behind the mug to grasp the handle. You find that the handle has an unusual shape, protruding outward to a sharp point. Plausibly, visual information about the cylinder and haptic information about the handle are combined in your overall shape experience. You experience the mug as a multi-part object consisting of a pointy handle side-attached to a cylinder. It is also plausible that your experience represents this shape in a view-invariant manner. You experience the handle as *side-attached to the cylinder*, not merely as located somewhere relative to your hand or eyes. Thus, the phenomenology in this case likely resides at Tier 3.

If this is correct, then Tier 3 shape phenomenology is constitutively multisensory under the unisensory-correspondence conception as well as the distinctiveness conception. The phenomenology of experiencing the mug's global shape could not be exhibited by any corresponding merely visual experience (given equivalent retinal stimulation but without haptic input) or merely haptic experience (vice versa). Either experience would inevitably leave out aspects of the mug's shape that are represented in the multisensory shape experience you actually enjoy.[xx]

5.3. Summing Up

This completes my solution to the cross-modal puzzle. When you see the pencil and then touch it, your experiences differ in salient respects. Thanks to Tier 1 phenomenology, you are aware of the pencil's "perspectival" properties, and these differ cross-modally. In vision, you are aware of the solid angle it subtends. In touch, you are aware of it as applying pressure, directly or indirectly, to certain body parts. Thanks to Tier 2 phenomenology, you are aware of the pencil's shape features in perceiver-centered reference frames that differ across modalities. However, thanks to Tier 3 phenomenology, you are aware of the very same geometrical features, specified in the very same way, through both modalities. At this tier, visual and haptic shape phenomenology are the same. But there is no *conflict* among the tiers, within or across modalities. Shape phenomenology is complex, not inconsistent. Its complexity is what allows us to resolve the cross-modal puzzle.

6. Imaginative Entanglement

I've argued that certain aspects of shape phenomenology are common to sight and touch. However, some philosophers have explicitly rejected this view. Recall Lopes' (2000) suggestion that visual and haptic experiences have "distinctive phenomenal characters through and through" (445; see also Block, 1996; O'Dea, 2006). While such claims are sometimes advanced without argument, the literature does contain a powerful challenge to the claim that aspects of spatial phenomenology are shared across modalities. I'll call this the *challenge from imaginative subtraction*.

The challenge from imaginative subtraction is usually raised as a problem for representationalism about perceptual experience. Representationalists hold that the phenomenal character of experience is determined by its representational content: Necessarily, any two experiences that differ in phenomenal character also differ in representational content (Tye, 2000, 69). Representationalism is often associated with the claim that spatial phenomenology is shared between experiences of different modalities. For, plausibly, certain spatial contents are shared across modalities. If so, and if spatial content fixes spatial phenomenal character, then some spatial phenomenal character should also be shared across modalities (cp. Lopes, 2000, 449). While representationalists can of course grant that visual and haptic experiences differ in many ways (e.g., we represent color in vision but not in touch, so visual experience has color phenomenology while haptic experience doesn't), the crucial point is that in respect of their shared *spatial contents*, the view seems to predict shared *spatial phenomenology*.[xxi]

Block (1996) challenges the idea that spatial phenomenology is shared between vision and audition. Although this is not the specific case that concerns us, Block's challenge generalizes to vision and touch (see O'Dea, 2006). Block argues that visual and auditory experience don't differ merely with respect to properties one modality represents and the other does not. They also differ in the phenomenal character associated with representing the very same property. Thus:

> Imagine the experience of hearing something and seeing it in your peripheral vision. It is true that you experience the sound as having a certain loudness, but can't we abstract away from that, concentrating on the perceived location? And isn't there an obvious difference between the auditory experience *as of that location* and the visual experience *as of that location*? (Block, 1996, 38)

Although visual and auditory experiences obviously differ thanks to our awareness of proper sensibles (e.g., color in vision, loudness in hearing), the idea is that we can *imaginatively subtract* the phenomenology associated with these proper sensibles to focus solely on our awareness of location. When we do so, we allegedly still find a phenomenal difference across modalities.

The same could be claimed for shape in sight and touch. Above I argued that vision and touch afford awareness of the same part-based, view-invariant structure. So some shape content is shared cross-modally. If representationalism is correct, then the phenomenology associated with this content must be shared cross-modally as well. But now a version of Block's challenge arises. When we focus on our visual and haptic experiences of a coffee mug, can't we imaginatively subtract factors like color phenomenology, thermal phenomenology, etc., to focus solely on the phenomenal aspect allegedly shared cross-modally—namely, our awareness of the mug's part-based, view-invariant structure? And when we do this, aren't we always left with some cross-modal difference? If so, then Tier 3 phenomenology *cannot* be shared across modalities.

Faced with this challenge, two options present themselves: First, grant that imaginative subtraction is possible (we can imaginatively isolate the phenomenal character shared across modalities), but reject the claim that cross-modal differences remain after subtraction is complete. Second, argue that imaginative subtraction is *not* possible for the phenomenal character under consideration, given our psychological constraints. Perhaps, try as you might, you *cannot* imaginatively isolate Tier 3 phenomenology.

I favor the second option. I propose that Tier 3 phenomenology is *imaginatively entangled* with other aspects of an experience's phenomenal character, some of which are modality-specific. Thus, whenever we attempt imaginative subtraction, cross-modal phenomenal differences remain. I don't believe that Tier 3 shape phenomenology is imaginatively entangled with *all* other aspects of phenomenal character. But I think that it is at least entangled with *Tier 1* phenomenology. We cannot perceptually imagine some detailed part-based, view-invariant structure without *also* imagining some perspectival features or other. When you visually imagine the view-invariant shape of a coffee mug, you are forced to imagine it from some perspective—as subtending some angle relative to your viewpoint. And in haptic imagery, you are forced to imagine exploring it with certain body parts, or via an intervening medium in contact with certain body parts. These perspectival properties needn't be imagined with maximal determinacy, but they cannot be wholly eliminated from an imaginative episode without eliminating view-invariant shape as well.[xxii]

Some remarks on imaginative entanglement. It's plausible that some features cannot be perceptually imagined in isolation. This is almost certainly true if we can perceptually imagine high-level properties like causation or animacy. We are unable to imagine one object *launching* another without imagining some more specific features of the event (e.g., rough direction of motion). Similar remarks likely hold for highly determinable features like topological closure.[xxiii]

Why suppose that the sensory representation of these features has any phenomenology at all, if we are unable to imagine that phenomenology in isolation? One reason is that awareness of high-level or determinable features governs phenomenal similarity relations. Two events that differ equally in their low-level motion properties are more different at the level of perceptual experience if they also differ in whether they exhibit launching (Kominsky et al., 2017). Likewise, two shapes that differ equally in their local features are more different at the level of experience if they also differ topologically (Todd et al., 2014; Green, 2017). The phenomenology of topology is real, but it cannot just float free—even in imagery—from the phenomenology of more specific shape properties. Imaginative subtraction will always fail for something like topology.

I now describe an empirically realistic view of sensory imagination on which Tier 3 phenomenology is imaginatively entangled with Tier 1 phenomenology. Here is the idea: Whenever we visually or haptically imagine an object's view-invariant shape in fine detail, we are also compelled to generate *array representations* that *obligatorily encode* perspectival features whenever they encode anything. The use of array representations during sensory imagery of shape is simply built into cognitive architecture, and such representations bring perspectival features along for the ride. Because the phenomenology associated with perspectival features is modality-specific, these modality-specific aspects are inevitably preserved during imagery.

Start with vision. It is widely accepted that vivid visual imagery recruits early visual cortex, which preserves the rough topography of the retina. This retinotopic organization is present during imagery (Slotnick et al., 2005; Kosslyn & Thompson, 2003; Pearson, 2019). Many theorists suggest that early visual cortical areas implement array-like representations structured in accordance with retinal topography. This dovetails with classic findings taken to support an array-based or depictive account of imagery, such as mental rotation and scanning (reviewed by Kosslyn et al., 2006). Consistent with the current proposal, Kosslyn and Thompson (2003) hold that visual imagery for detailed shape obligatorily recruits arrays in early visual cortex.

Recall that on one account (Tye, 1991; Lande, 2018), visual arrays are composed of cells each dedicated to a particular direction or line of sight. Adjacent cells correspond to adjacent directions (which project to adjacent regions of the retina), and symbols within cells encode features like distance, orientation, or color. Suppose, then, that visual imagery of shape always recruits array representations of this sort. Such arrays obligatorily encode direction: An object cannot be represented in visual array format without representing something about the directions to its boundary points (though direction might be represented with less than absolute precision using cells that pool over a range of directions). The directions to an object's boundary points determine its solid angle. If detailed visual shape imagery always recruits array representations, then plausibly it always represents distinctively visual perspectival features like solid angle.[xxiv]

Turn to touch. While haptic imagery is less studied than vision, it appears to recruit primary somatosensory cortex, where adjacent neural structures correspond to adjacent body parts or skin regions (Schmidt et al., 2014; de Borst & de Gelder, 2017). Somatotopic activity during tactile imagery matches imagery content: Imagining a stimulus on the right hand activates regions of somatosensory cortex dedicated to that hand (Schmidt & Blankenburg, 2019). A *tactile array*, let's say, is composed of cells each dedicated to a particular region of the body or skin. Adjacent cells map to adjacent bodily regions, and symbols within cells encode information about the magnitude of pressure an object applies to that body part, or the texture/temperature of the stimulus at that somatotopic location.[xxv]

To represent an object via tactile array, one must encode at least its approximate somatotopic projection. An object can't be represented without taking some stance on where it pressures the body, though again this might be represented with some imprecision. If haptic imagery of shape mandates the use of tactile arrays in somatosensory cortex (Schmidt & Blankenburg, 2019, 6), then it represents the sorts of features involved in Tier 1 haptic phenomenology, whose representation is distinctive to touch. Haptic shape imagery may require more besides tactile arrays (e.g., motor imagery of exploratory movement), but that does not challenge the current proposal.[xxvi]

While this model is admittedly speculative, it would explain why Tier 3 phenomenology is imaginatively entangled with Tier 1 phenomenology in both vision and touch. Furthermore, it allows that Tier 3 phenomenology is entangled with Tier 1 phenomenology but *not* with color or thermal phenomenology. A visual-array cell might represent a surface patch of such-and-such distance in such-and-such direction, but remain silent about color (Tye, 1991, 93). A tactile-array cell might represent the presence of an object applying pressure at a given bodily location while remaining silent about temperature or texture. The present view thus permits that we may imaginatively isolate shape phenomenology from color or thermal phenomenology.

I propose that Tier 3 phenomenology plays a similar role in perceptual experience to the phenomenology of high-level or determinable features. While these aspects of phenomenal character cannot stand alone in perceptual experience, they are introspectively detectable in patterns of phenomenal similarity. Tier 3 phenomenology explains why the visual experience of a wrench is more like the haptic experience of a wrench than that of a screwdriver. These shared phenomenal features underlie our ability to compare shapes across modalities in both shape resemblance judgments and cross-modal recognition.

This reply to the challenge from imaginative subtraction is congenial to representationalism. Representationalists shouldn't be embarrassed by the commitment to cross-modally shared shape phenomenology, since that commitment is empirically supported and defensible against objections. This falls short of *vindicating* representationalism, even in the restricted domain of spatial experience. To do that, other challenges need answers (Masrour, 2017). Moreover, as flagged in section 3.1, it remains controversial whether representationalists can explain Tier 1 visual phenomenology. However, these issues must be left for another time.

7. Conclusion

The *puzzle of cross-modal shape experience* is the puzzle of reconciling the apparent differences between visual and haptic shape experiences with their apparent similarities. I've argued that we can resolve the cross-modal puzzle by reflecting on the *puzzle of perspectival character*, which challenges us to reconcile the apparent variability of shape experience with shifts in perspective with its apparent constancy. On the model of shape phenomenology I've defended, visual and haptic experiences differ in their perspectival aspects, but share their constant aspects. I've assembled both empirical and introspective evidence for this view. I've also considered a powerful challenge to the view that visual and haptic experiences share elements of shape phenomenology. I've argued that the challenge can be answered by appeal to constraints on the architecture of sensory imagination.[xxvii]

References

Amedi, A., Malach, R., Hendler, T., Peled, S., & Zohary, E. (2001). Visuo-haptic object-related activation in the ventral visual pathway. *Nature Neuroscience*, *4*(3), 324-330.

Amedi, A., Jacobson, G., Hendler, T., Malach, R., & Zohary, E. (2002). Convergence of visual and tactile shape processing in the human lateral occipital complex. *Cerebral Cortex*, *1*, 1202-1212.

Amedi, A., Raz, N., Azulay, H., Malach, R., & Zohary, E. (2010). Cortical activity during tactile exploration of objects in blind and sighted humans. *Restorative Neurology and Neuroscience*, *28*(2), 143-156.

Ayzenberg, V., Chen, Y., Yousif, S. R., & Lourenco, S. F. (2019). Skeletal representations of shape in human vision: Evidence for a pruned medial axis model. *Journal of vision*, *19*(6):6, 1-21.

Azañón, E., Longo, M. R., Soto-Faraco, S., & Haggard, P. (2010). The posterior parietal cortex remaps touch into external space. *Current Biology*, *20*(14), 1304-1309.

Azañón, E., & Soto-Faraco, S. (2008). Changing reference frames during the encoding of tactile events. *Current Biology*, *18*(14), 1044-1049.

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.

Bayne, T. (2010). *The Unity of Consciousness*. Oxford: Oxford University Press.

Bennett, D. J. (2012). Seeing shape: Shape appearances and shape constancy. *The British Journal for the Philosophy of Science*, *63*(3), 487-518.

Bennett, D. J., & Vuong, Q. C. (2006). A stereo advantage in generalizing over changes in viewpoint on object recognition tasks. *Perception & Psychophysics*, *68*(7), 1082-1093.

Berkeley, G. (1709/1965). *Essay Towards a New Theory of Vision*. In D. M. Armstrong (ed.), *Berkeley's Philosophical Writings*. New York: Collier Books.

Bermúdez, J.-L. (2005). The phenomenology of bodily awareness. In L. Thomasson, & D. W. Smith (eds.), *Phenomenology and Philosophy of Mind* (pp. 259-317). Oxford: Oxford University Press.

Berryman, L. J., Yau, J. M., & Hsiao, S. S. (2006). Representation of object size in the somatosensory system. *Journal of Neurophysiology*, *96*(1), 27-39.

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

Blazhenkova, O. (2016). Vividness of object and spatial imagery. *Perceptual and Motor Skills*, *122*(2), 490-508.

Block, N. (1983). The photographic fallacy in the debate about mental imagery. *Nous*, *17*(4), 651-661.

Block, N. (1996). Mental paint and mental latex. *Philosophical Issues*, *7*, 19-49.

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

Bourget, D. (2017). Representationalism and sensory modalities: An argument for intermodal representationalism. *American Philosophical Quarterly*, *54*, 251-268.

Brewer, B. (2011). *Perception and its Objects*. Oxford: Oxford University Press.

Briscoe, R. E. (2008). Vision, action, and make-perceive. *Mind & Language*, *23*(4), 457-497.

Briscoe, R. E. (2019). Bodily awareness and novel multisensory features. *Synthese*. DOI: 10.1007/s11229-019-02156-2.

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

Burge, T. (2014). Reply to Rescorla and Peacocke: Perceptual content in light of perceptual constancies and biological constraints. *Philosophy and Phenomenological Research*, *88*(2), 485-501.

Byrne, A., & Hilbert, D. R. (2003). Color realism and color science. *Behavioral and Brain Sciences*, *26*(1), 3-21.

Byrne, A., & Tye, M. (2006). Qualia ain't in the head. *Noûs*, *40*(2), 241-255.

Campbell, J. (2002). *Reference and Consciousness*. Oxford: Oxford University Press.

Chalmers, D. J. (1996). *The Conscious Mind: In Search of a Fundamental Theory*. Oxford: Oxford University Press.

Cohen, E. H., & Singh, M. (2006). Perceived orientation of complex shape reflects graded part decomposition. *Journal of Vision*, *6*, 805-821.

Cohen, J. (2010). Perception and computation. *Philosophical Issues*, *20*, 96-124.

Cristino, F., Davitt, L., Hayward, W. G., & Leek, E. C. (2015). Stereo disparity facilitates view generalization during shape recognition for solid multipart objects. *Quarterly Journal of Experimental Psychology*, *68*(12), 2419-2436.

de Borst, A. W., & de Gelder, B. (2017). fMRI-based multivariate pattern analyses reveal imagery modality and imagery content specific representations in primary somatosensory, motor and auditory cortices. *Cerebral Cortex*, *27*(8), 3994-4009.

de Vignemont, F., & Massin, O. (2015). Touch. In M. Matthen (ed.), *The Oxford Handbook of Philosophy of Perception*. Oxford: Oxford University Press.

Dretske, F. I. (1995). *Naturalizing the Mind*. Cambridge, MA: MIT Press.

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, *96*(3), 433-458.

Erdogan, G., Chen, Q., Garcea, F. E., Mahon, B. Z., & Jacobs, R. A. (2016). Multisensory part-based representations of objects in human lateral occipital cortex. *Journal of Cognitive Neuroscience*, *28*(6), 869-881.

Erdogan, G., Yildirim, I., & Jacobs, R. A. (2015). From sensory signals to modality-independent conceptual representations: A probabilistic language of thought approach. *PLoS Computational Biology*, *11*, e1004610.

Evans, G. (1985). Molyneux's question. In A. Phillips (ed.), *Gareth Evans: Collected Papers*. Oxford: Clarendon Press.

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

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*, *25*(2), 377-386.

Fleming, R. W., Torralba, A., & Adelson, E. H. (2004). Specular reflections and the perception of shape. *Journal of Vision*, *4*, 798-820.

French, C. (2018). Bálint's syndrome, object seeing, and spatial perception. *Mind & Language*, *33*(3), 221-241.

Friedman-Hill, S. R., Robertson, L. C., & Treisman, A. (1995). Parietal contributions to visual feature binding: Evidence from a patient with bilateral lesions. *Science*, *269*(5225), 853-855.

Fulkerson, M. (2014). *The First Sense: A Philosophical Study of Human Touch.* Cambridge, MA: MIT

Press.

Gaißert, N., Wallraven, C., & Bülthoff, H. H. (2010). Visual and haptic perceptual spaces show high similarity in humans. *Journal of Vision*, *10*(11):2, 1-20.

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, *15*, 20-25.

Green, E. J. (2017). A layered view of shape perception. *The British Journal for the Philosophy of Science*, *68*(2), 255-387.

Green, E. J. (2019). On the perception of structure. *Noûs*, *53*(3), 564-592.

Green, E. J. (2020). Representing shape in sight and touch. *Mind & Language*. DOI: 10.1111/mila.12352.

Green, E. J., & Schellenberg, S. (2018). Spatial perception: The perspectival aspect of perception. *Philosophy Compass*, *13*(2), e12472.

Haggard, P., & Giovagnoli, G. (2011). Spatial patterns in tactile perception: is there a tactile field? *Acta Psychologica*, *137*(1), 65-75.

Harrar, V., & Harris, L. R. (2010). Touch used to guide action is partially coded in a visual reference

frame. *Experimental Brain Research*, *203*(3), 615-620.

Harrison, S. J., & Feldman, J. (2009). The influence of shape and skeletal axis structure on texture perception. *Journal of Vision*, *9*(6):13, 1-21.

Hill, C. S. (1991). *Sensations: A Defense of Type Materialism*. Cambridge: Cambridge University Press.

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

Hill, C. S. (2016). Perceptual relativity. *Philosophical Topics*, *44*(2), 179-200.

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

Hopp, W. (2013). No such look: Problems with the dual content theory. *Phenomenology and the Cognitive Sciences*, *12*(4), 813-833.

Huemer, M. (2001). *Skepticism and the Veil of Perception*. Lanham, MD: Rowman & Littlefield.

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

Hung, C. C., Carlson, E. T., & Connor, C. E. (2012). Medial axis shape coding in macaque inferotemporal cortex. *Neuron*, *74*(6), 1099-1113.

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

Jagnow, R. (2012). Representationalism and the perspectival character of perceptual experience. *Philosophical Studies*, *157*(2), 227-249.

James, T. W., Humphrey, G. K., Gati, J. S., Menon, R. S., & Goodale, M. A. (2002). Differential effects of viewpoint on object-driven activation in dorsal and ventral streams. *Neuron*, *35*(4), 793-801.

James, T. W., James, K. H., Humphrey, G. K., & Goodale, M. A. (2006). Do visual and tactile object representations share the same neural substrate? In M. A. Heller & S. Ballesteros (eds.), *Touch and Blindness: Psychology and Neuroscience*. Mahwah, NJ: Lawrence Erlbaum, 139-155.

Kappers, A. M., & Viergever, R. F. (2006). Hand orientation is insufficiently compensated for in haptic spatial perception. *Experimental Brain Research*, *173*(3), 407-414.

Keogh, R., & Pearson, J. (2018). The blind mind: No sensory visual imagery in aphantasia. *Cortex*, *105*, 53-60.

Kim, S. S., Gomez-Ramirez, M., Thakur, P. H., & Hsiao, S. S. (2015). Multimodal interactions between proprioceptive and cutaneous signals in primary somatosensory cortex. *Neuron*, *86*(2), 555-566.

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

Kosslyn, S. M., & Thompson, W. L. (2003). When is early visual cortex activated during visual mental imagery? *Psychological Bulletin*, *129*(5), 723-746.

Kosslyn, S. M., Thompson, W. L., & Ganis, G. (2006). *The Case for Mental Imagery*. Oxford: Oxford University Press.

Kovács, I., Fehér, A., & Julesz, B. (1998). Medial-point description of shape: A representation for action coding and its psychophysical correlates. *Vision Research*, *38*(15-16), 2323-2333.

Lacey, S., Hall, J., & Sathian, K. (2010). Are surface properties integrated into visuohaptic object representations?. *European Journal of Neuroscience*, *31*(10), 1882-1888.

Lacey, S., Pappas, M., Kreps, A., Lee, K., & Sathian, K. (2009). Perceptual learning of view-independence in visuo-haptic object representations. *Experimental Brain Research*, *198*(2-3), 329-337.

Lacey, S., Peters, A., & Sathian, K. (2007). Cross-modal object recognition is viewpoint-independent. *PLoS One*, *2*(9), e890.

Lacey, S., & Sathian, K. (2014). Visuo-haptic multisensory object recognition, categorization, and representation. *Frontiers in Psychology*, *5*, 730.

Lande, K. J. (2018). The perspectival character of perception. *The Journal of Philosophy*, *115*(4), 187-214.

Lawson, R. (2009). A comparison of the effects of depth rotation on visual and haptic three-dimensional object recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(4), 911-930.

Lawson, R. (2011). An investigation into the cause of orientation-sensitivity in haptic object recognition. *Seeing and Perceiving*, *24*(3), 293-314.

Lawson, R., Boylan, A., & Edwards, L. (2014). Where you look can influence haptic object recognition. *Attention, Perception, & Psychophysics*, *76*(2), 559-574.

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

Levin, J. (2018). Molyneux's question and the amodality of spatial experience. *Inquiry*, *61*(5-6), 590-610.

Linton, P. (2021). Conflicting shape percepts explained by perception cognition distinction. *Proceedings of the National Academy of Sciences*, *118*(10), e2024195118.

Locke, J. (1694/1975). *An Essay Concerning Human Understanding*. P. H. Nidditch (ed.). Oxford: Oxford University Press.

Lopes, D. M. M. (2000). What is it like to see with your ears?: The representational theory of mind. *Philosophy and Phenomenological Research*, *60*(2), 439-453.

Lowet, A. S., Firestone, C., & Scholl, B. J. (2018). Seeing structure: Shape skeletons modulate perceived similarity. *Attention, Perception, & Psychophysics*, *80*(5), 1278-1289.

Lycan, W. (1996). *Consciousness and Experience*. Cambridge, MA: MIT Press.

Mac Cumhaill, C. (2017). The tactual ground, immersion, and the "space between." *The Southern Journal of Philosophy*, *55*(1), 5-31.

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.

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

Martin, M. (1992). Sight and touch. In T. Crane (ed.), *The Contents of Experience*. Cambridge: Cambridge University Press, 196-215.

Masrour, F. (2017). Space perception, visual dissonance and the fate of standard representationalism. *Noûs*, *51*(3), 565-593.

Masson, H. L., Bulthé, J., Op de Beeck, H. P., & Wallraven, C. (2016). Visual and haptic shape processing in the human brain: unisensory processing, multisensory convergence, and top-down influences. *Cerebral Cortex*, *26*(8), 3402-3412.

Matthen, M. (forthcoming). The dual structure of touch: The body vs. peripersonal space. In F. de Vignemont, A. Serino, H. Y. Wong, & A. Farné (eds.), *The World at Our Fingertips*. Oxford: Oxford University Press.

Matthen, M., & Cohen, J. (2020). Many Molyneux questions. *Australasian Journal of Philosophy*, *98*(1), 47-63.

Morales, J., Bax, A., & Firestone, C. (2020). Sustained representation of perspectival shape. *Proceedings of the National Academy of Sciences*, *117*(26), 14873-14882.

Morales, J., Bax, A., & Firestone, C. (ms). Perspectival interference up close.

Munton, J. (2020). Visual indeterminacy and the puzzle of the speckled hen. *Mind & Language*. DOI: 10.1111/mila.12296.

Newell, F. N., Ernst, M. O., Tjan, B. S., & Bülthoff, H. H. (2001). Viewpoint dependence in visual and haptic object recognition. *Psychological Science*, *12*(1), 37-42.

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

O'Callaghan, C. (2015). The multisensory character of perception. *The Journal of Philosophy*, *112*(10), 551-569.

O'Callaghan, C. (2019). *A Multisensory Philosophy of Perception*. Oxford: Oxford University Press.

O'Dea, J. W. (2006). Representationalism, supervenience, and the cross-modal problem. *Philosophical Studies*, *130*(2), 285-295.

O'Shaughnessy, B. (1989). The sense of touch. *Australasian Journal of Philosophy*, *67*(1), 37-58.

Pautz, A. (2017). Experiences are representations: An empirical argument. In B. Nanay (ed.), *Current Debates in Philosophy of Perception* (pp. 23-42). New York: Routledge.

Pautz, A. (2020). The puzzle of the laws of appearance. *Philosophical Issues*, *30*(1), 257-272.

Peacocke, C. (1983). *Sense and Content: Experience, Thought and their Relations.* Oxford: Oxford University Press.

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

Pearson, J. (2019). The human imagination: The cognitive neuroscience of visual mental imagery. *Nature Reviews Neuroscience*, *20*, 624-634.

Peissig, J. J., & Tarr, M. J. (2007). Visual object recognition: Do we know more now than we did 20 years ago? *Annual Review of Psychology*, *58*, 75-96.

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

Poggio, T., & Edelman, S. (1990). A network that learns to recognize three-dimensional objects. *Nature*, *343*(6255), 263-266.

Prinz, J. (2002). *Furnishing the Mind: Concepts and Their Perceptual Basis*. Cambridge, MA: MIT Press.

Rauschenberger, R., & Yantis, S. (2001). Masking unveils pre-amodal completion representation in visual search. *Nature*, *410*(6826), 369-372.

Richardson, L. (2014). Space, Time and Molyneux's Question. *Ratio*, *27*(4), 483-505.

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

Schmidt, T. T., & Blankenburg, F. (2019). The somatotopy of mental tactile imagery. *Frontiers in Human Neuroscience*, *13*(10), 1-10.

Schmidt, T. T., Ostwald, D., & Blankenburg, F. (2014). Imaging tactile imagery: changes in brain connectivity support perceptual grounding of mental images in primary sensory cortices. *Neuroimage*, *98*, 216-224.

Schwenkler, J. (2012). Does visual spatial awareness require the visual awareness of space? *Mind & Language*, *27*(3), 308-329.

Schwenkler, J. (2013). Do things look the way they feel? *Analysis*, *73*(1), 86-96.

Schwenkler, J. (2019). Molyneux's question within and across the senses. In T. Cheng, O. Deroy, & C. Spence (eds.), *Spatial Senses: Philosophy of Perception in an Age of Science*. London: Routledge, 274-283.

Schwitzgebel, E. (2006). Do things look flat? *Philosophy and Phenomenological Research*, *72*(3), 589-599.

Schwitzgebel, E. (2011). *Perplexities of Consciousness*. Cambridge, MA: MIT 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. *Advances in Psychology*, *130*, 401-459.

Slotnick, S. D., Thompson, W. L., & Kosslyn, S. M. (2005). Visual mental imagery induces retinotopically organized activation of early visual areas. *Cerebral Cortex*, *15*(10), 1570-1583.

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

Stazicker, J. (2018). The visual presence of determinable properties. In F. Dorsch & F. Macpherson (eds.), *Phenomenal Presence* (pp. 105-133). Oxford: Oxford University Press.

Tal, N., & Amedi, A. (2009). Multisensory visual–tactile object related network in humans: insights gained using a novel crossmodal adaptation approach. *Experimental Brain Research*, *198*(2-3), 165-182.

Tarr, M. J., & Pinker, S. (1989). Mental rotation and orientation-dependence in shape recognition. *Cognitive Psychology*, *21*(2), 233-282.

Todd, J. T., Weismantel, E., & Kallie, C. S. (2014). On the relative detectability of configural properties. *Journal of Vision*, *14*(1):18, 1-8.

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

Tye, M. (2000). *Consciousness, Color, and Content*. Cambridge, MA: MIT Press.

Ueda, Y., & Saiki, J. (2012). Characteristics of eye movements in 3-D object learning: comparison between within-modal and cross-modal object recognition. *Perception*, *41*(11), 1289-1298.

van der Horst, B. J., & Kappers, A. M. (2008). Using curvature information in haptic shape perception of 3D objects. *Experimental Brain Research*, *190*(3), 361-367.

van Mier, H. I. (2014). Haptic perception of parallelity. *Psychology and Behavioral Sciences*, *3*(6), 212-221.

Volcic, R., & Kappers, A. M. (2008). Allocentric and egocentric reference frames in the processing of three-dimensional haptic space. *Experimental Brain Research*, *188*(2), 199-213.

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

Yamamoto, S., & Kitazawa, S. (2001). Reversal of subjective temporal order due to arm crossing. *Nature Neuroscience*, *4*(7), 759-765.

Yau, J. M., Kim, S. S., Thakur, P. H., & Bensmaia, S. J. (2016). Feeling form: the neural basis of haptic shape perception. *Journal of Neurophysiology*, *115*(2), 631-642.

Zanker, J. M., & Quenzer, T. (1999). How to tell circles from ellipses: perceiving the regularity of simple shapes. *Naturwissenschaften*, *86*(10), 492-495.

---

[i] See also Block (1996), Prinz (2002, ch. 5), and O'Dea (2006). Berkeley (1709/1965) is an important precursor.
[ii] O'Callaghan expresses sympathy with this view but is not strongly committed to it. Dretske (1995, 95), Bayne (2010, 62-64), and Bourget (2017) are so committed.
[iii] Bayne (2010, 63) floats this idea but also holds that visual and haptic experiences differ in other ways.
[iv] Martin (1992, 209-210) articulates a similar idea (see also O'Shaughnessy, 1989, 38).
[v] Reply: Perhaps Bálint's patients *do* visually experience objects as occupying regions of a spatial field, but the field shrinks to contain just one object and nothing else (French, 2018). However, if our conception of the visual field permits "tunnel vision" of this sort, then I remain doubtful that it grounds the distinction between visual and haptic shape experience. When we feel a cube, perhaps we are not aware of a broader region of space beyond the cube where no objects are felt (although see Mac Cumhaill, 2017), but don't we at least become aware of the region the cube occupies? If so, then visual awareness of a spatial field confined to a single object is not an attractive way to distinguish visual and haptic shape experience, since something similar occurs in touch.
[vi] Pizlo (2008, ch. 1) criticizes the overreliance on elliptical stimuli in shape constancy experiments.
[vii] Another objection might be that the slanted circles were *misperceived* as elliptical in the Morales et al. study (e.g., if subjects displayed underconstancy (Linton, 2021)). Morales et al. try to preempt this worry by pointing out that identification of the head-on ellipse was nearly at ceiling (exp. 1). However, future experiments should more fully investigate the extent to which misperception might have occurred (some such work is already underway—see

Morales et al. ms.). I regard the Morales et al. results as *suggestive* evidence for perspectival similarity that *complement* commonplace introspective observations. The findings are not conclusive on their own.

[viii] A related view holds that experience represents an object's *P-shape*, or the 2D shape needed to occlude the object on a plane perpendicular to the line of sight (Noë ,2004). There are reasons to prefer the solid-angle view to the P-shape view. Notably, the former view does not entail that when we visually experience a volumetric object we perceptually represent some *flat* apparent shape, which seems implausible on introspective grounds (Schwitzgebel, 2006).

[ix] A similar point holds for Hill's (2014, 2016) proposal that perspectival variation and similarity are explained by our perceptual awareness of *Thouless properties*, which are the outputs of imperfect constancy operations. A slanted coin's Thouless shape is compressed relative to its real shape, but not *as* compressed as its retinal image. On this view, Tier 1 phenomenology is still conditioned by the retinal image. The retinal image is the original *source* of the patterns of compression associated with slant. If the proximal stimulus for vision did not distort in this way, there would be no reason to expect imperfect shape constancy operations to yield a compressed Thouless shape.

[x] One might doubt that genuine shape experience is involved in distal touch. Perhaps you experience the *presence of an object* without experiencing its shape. I disagree, but if so, then the analysis of Tier 1 haptic phenomenology can be modified to rule out distal touch by restricting it to awareness of *direct* pressure to the body. It might also be claimed that when you explore a rock with a stick, your inner body map actually expands to encompass the stick (Martin, 1992, 201-202; de Vignemont & Massin, 2015), so the object *is* experienced as directly contacting the body after all. But plausibly this is not *always* the case (Fulkerson, 2014, 145-146); only a single counterexample is needed to motivate the provision for indirect application of pressure.

[xi] Thanks to an anonymous reviewer for this suggestion.

[xii] These include shading, texture gradients, and specular reflections. Such cues also remain unchanged through mere changes in which retina an object stimulates during monocular viewing.

[xiii] Briscoe (2008) discusses these images in detail.

[xiv] These representations may call to mind Marr's (1982) 2½-D sketch. However, they are not *exactly* like the 2½-D sketch, since the 2½-D does not explicitly specify boundaries between discrete objects and surfaces. To support recognition, viewer-centered information about depth and orientation must be grouped or segmented in order to specify which collections of surface patches belong to an individual object (Jackendoff, 1987, 331-338).

[xv] Haptic parallelity judgments are, however, not *solely* determined by hand-centered coordinates. They incorporate a weighted average of hand-centered and allocentric reference frames (Volcic & Kappers, 2008).

[xvi] I should note that there is evidence that haptic recognition can also be sensitive to head position. Under certain conditions (viz., when reaching one's hand across the body midline), objects are more easily recognized when the observer looks in their direction, even if the objects aren't visible (Lawson et al., 2014). Thus, while perceiver-centered haptic shape representation plausibly *prioritizes* hand-centered reference frames during manual exploration, head-centered reference frames may be involved as well, at least under some circumstances. One hypothesis is that perceiver-centered haptic shape representation employs a mixed coordinate system involving, say, a weighted average of head-centered and hand-centered reference frames (Harrar & Harris, 2010, 618), with greater weight assigned to the latter. Nonetheless, as long as Tier 2 haptic phenomenology is *predominantly* determined by frames centered on the body parts used for exploration, it is reasonable to conclude that Tier 2 phenomenology differs cross-modally (see below).

[xvii] Proponents of structural schemes include Marr and Nishihara (1978), Biederman (1987), Hummel (2013), and Green (2019).

[xviii] It might be suggested that LOC activation during haptic exploration reflects visual imagery rather than haptic shape processing. Against this, Amedi et al. (2001) found greater LOC activation during haptic shape perception than during visual imagery alone, and Amedi et al. (2010) found LOC recruitment during haptic shape perception in congenitally blind subjects. However, it is possible that visual imagery is more involved in the haptic processing of familiar shapes than unfamiliar shapes (Lacey & Sathian, 2014).

[xix] Likewise for mixed views on which phenomenology is determined partly by inner physical-functional properties and partly by the world (Lycan, 1996), since both elements of the determination base are shared cross-modally.

[xx] In O'Callaghan's (2019, 67-77) taxonomy, the mug's shape is a novel multisensory feature *instance*, but not a novel multisensory feature *type*. It is a feature instance that can only be experienced through the coordinated operation of multiple modalities. However, other instances of the same feature type (viz., shape) can be experienced without multisensory coordination.

[xxi] Strictly speaking, representationalism does not require that *spatial* content fix *spatial* phenomenal character, only that *overall* content fix *overall* phenomenal character. Spatial phenomenology could be determined jointly by spatial

and non-spatial content. However, in practice most representationalists have endorsed the local determination of aspects of phenomenology by corresponding aspects of content. For instance, it is held that the color phenomenology is fixed by the representation of surface spectral reflectance (Byrne & Hilbert, 2003). Such representationalists would presumably endorse the analogous claim that spatial content fixes spatial phenomenology.

[xxii] Pautz (2020) endorses similar claims. Pautz contends that it is a "law of appearance" that nothing can appear to have a particular intrinsic shape without appearing to have some perspectival "shape from here." Two differences: First, Pautz's primary concern is perceptual experience, while mine is sensory imagery. Second, Pautz holds that the laws of appearance are metaphysically necessary. For my purposes, it's enough that relations of imaginative entanglement arise from contingent special science laws.

[xxiii] I do not endorse the more extreme claim that sensory imagery of a dimension requires *maximal* determinacy along that dimension. Experience sometimes represents determinables without any corresponding determinate (Block, 1983; Stazicker, 2018; Munton, 2020).

[xxiv] I do not claim that *all* visual imagery obligatorily recruits array representations—only the detailed imagery of shape. There is evidence for a distinction between *object* imagery (involving rich features like shape, color, and texture) and *spatial* imagery (involving schematic positions and spatial arrangements of objects/scenes) (Blazhenkova, 2016). It has been suggested that depictive arrays in early visual cortex are recruited only for object imagery (Keogh & Pearson 2018).

[xxv] Compare the familiar notion of a "body map" (de Vignemont & Massin, 2015).

[xxvi] Tactile arrays might also combine information about pressure with information about posture in intricate ways. Recent findings indicate that a large proportion of neurons in primary somatosensory cortex are responsive to both cutaneous and proprioceptive inputs (Kim et al., 2015).

[xxvii] I am indebted to Chris Hill, Adam Pautz, Lilian Jin, David Bennett, Kevin Lande, Alex Byrne, Jake Quilty-Dunn, David Chalmers, Bence Nanay, Chris Gauker, and Becko Copenhaver for valuable discussion and feedback. I also benefited from presenting portions of this material at Uriah Kriegel's Autumn of Consciousness talk series, Chris Hill and Adam Pautz's Fall 2020 Brown graduate seminar, the WUSTL Mind and Perception Group, the Antwerp/Salzburg Imagistic Cognition Group, and the MIT Work-in-Progress seminar. Finally, thanks to two anonymous reviewers for their very helpful comments.