Representing shape in sight and touch

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.

| Citation       | Green, E. J. “Representing shape in sight and touch.” Mind and Language (December 2020) © 2020 The Author |
|----------------|-------------------------------------------------------------------------------------------------------|
| As Published   | 10.1111/mila.12352                                                                                       |
| Publisher      | Wiley                                                                                                     |
| Version        | Author’s final manuscript                                                                               |
| Citable link   | https://hdl.handle.net/1721.1/130161                                                                         |
| Terms of Use   | Creative Commons Attribution-Noncommercial-Share Alike                                                   |
| Detailed Terms | http://creativecommons.org/licenses/by-nc-sa/4.0/                                                        |
We represent shape in both sight and touch, but how do these abilities relate to one another? This issue has been discussed in the context of Molyneux’s question of whether someone born blind could, upon being granted sight, identify shapes visually. Some have suggested that we might look to real-world cases of sight restoration to illuminate the relation between visual and tactual shape representations. Here I argue that newly sighted perceivers should not be relied on in this way because they are unlikely to form the kinds of shape representations responsible for cross-modal recognition in normally sighted perceivers. I then argue that the available evidence makes a compelling case for the type identity view, on which the visual and tactual representations responsible for cross-modal recognition are type-identical.

KEYWORDS
shape perception, Molyneux’s question, multisensory perception, tactile perception, perception
1. INTRODUCTION

Some properties are perceptible only through a single modality, while others are perceptible through multiple modalities. Shape is a prototypical case of the latter. We perceptually represent spheres and cubes in both sight and touch. But how do our visual and tactual representations of shape relate to one another, and what could settle the issue?

Philosophical discussions of visual and tactual shape perception have generally been framed in the context of Molyneux’s question, which asks whether someone born blind could, upon being granted sight, recognize shapes visually. Recently, some have suggested that we might examine real-world cases of sight restoration to illuminate the relation between visual and tactual shape representations. Here I argue that newly sighted subjects should not be relied on for this purpose because they are unlikely to form the kinds of shape representations responsible for cross-modal recognition in normal perceivers. I then defend a positive view about the relation between visual and tactual shape representations. I argue that the visual and tactual representations directly responsible for cross-modal shape recognition are type-identical. This explains our ability to recognize previously touched shapes on the basis of sight, and vice versa.

2. MOLYNEUX’S QUESTION

In a letter to John Locke in 1693, William Molyneux invites us to consider a man born blind who has learned to recognize spheres and cubes by touch. Suppose he were suddenly granted sight.
“Quaere, whether by his sight, before he touched them, he could now distinguish, and tell, which is the Globe, which the Cube” (Locke, 1694/1975, Book II, Chapter IX, Section 8).

Molyneux’s question raises many issues. The question has been used to structure discussions of the relation between the shape concepts applied on the basis of vision and touch (Levin, 2008), the phenomenal character of shape experience (Campbell, 1996), and the issue of whether vision and touch employ a common spatial reference frame (Evans, 1985). It is possible that Molyneux himself intended to raise the issue of whether distance is an original object of sight (Degenaar, 1996, Chapter 2).¹ Here I focus on its connection to the issue of whether there is an intrinsic similarity between our visual and tactual representations of shape.

We recognize shapes through both vision and touch, and novel shapes first encountered through one modality can be reidentified through the other (Norman et al., 2004). What explains these abilities? Two answers present themselves. First, perhaps vision and touch produce intrinsically similar representations of shape. Because we are sensitive to this similarity, we are prepared to apply shape concepts on the basis of representations in either modality. Second, perhaps vision and touch produce intrinsically dissimilar representations of shape. If so, the ability to apply shape concepts on the basis of both modalities relies on learned or hardwired connections between these representations.

Molyneux’s question is supposed to help us decide which answer is correct. If cross-modal recognition relies on an intrinsic similarity between sight and touch, then presumably a newly sighted perceiver who is able to perceive shape through both sight and touch should have access to this similarity. Accordingly, if the perceiver encounters a shape through touch but fails to recognize it visually, this suggests that their visual and tactual shape representations are intrinsically dissimilar—or, more precisely, too dissimilar to ground application of the same

¹ For even broader generalizations of Molyneux’s question, see Matthen and Cohen (2020).
shape concepts. So if newly sighted subjects fail to recognize shapes across modalities, this is supposed to support the second answer: Shape representations in sight and touch are intrinsically dissimilar, and cross-modal recognition depends on learned connections between them.²

This is clearly the role Berkeley envisioned for the newly sighted subject. The rationale was to consider someone whose visual and tactual ideas are intrinsically like ours, but stripped of the effects of “custom, and erroneous suggestions of prejudice” (Berkeley, 1709/1965, Section 128). If this person fails to recognize shapes by means of sight, then we can conclude that our visual and tactual ideas of shape differ from one another. Berkeley, of course, endorsed a further move from this to the conclusion that the objects of sight and touch are different, but few would accept this argument today.

More recently, Schwenkler (2012, 2013) also understands Molyneux’s question to bear on the issue of an intrinsic similarity between visual and tactual shape representations. He writes:

[T]hink of the way that shape and other spatial properties are perceived in sight and touch, of how we can tell right away whether a seen shape is the same as some felt one. Molyneux’s question asks: can we do this only because of learned associations built up in the course of past experience, or are the representations of these properties related somehow intrinsically? (Schwenkler, 2012, p. 186)

Similarly, Hopkins (2005) takes the main issue raised by Molyneux’s question to concern whether vision and touch so represent shape “as to support different concepts of it” (p. 442). If our visual and tactual representations of shape differ intrinsically, then they may support

² Success in the task is harder to interpret, since it could be due either to intrinsic similarity between the shape representations or to hardwired connections between them (Evans, 1985).
different shape concepts. Molyneux’s test is supposed to provide a method for determining whether this is true.

I have appealed to the idea of an “intrinsic similarity” between the visual and tactual representations responsible for cross-modal recognition. What would such intrinsic similarity involve? There are three obvious possibilities.

First, it is possible that vision and touch generate tokens of the same representation type. Suppose F is a shape property perceptible through both sight and touch. Then, on this proposal, there is a sensory representation type R such that R represents F, and tokens of R are produced when instances of F are perceived either by sight alone or by touch alone. Call this the type identity view.

Second, it is possible that vision and touch generate distinct representation types that are nonetheless rationally linked. On this proposal, vision and touch represent shape by means of different representation types, but the contents of these representations are such that the subject cannot rationally doubt that they represent the same property. For example, they might employ different descriptions that are analytically equivalent (e.g., “closed four-sided figure” vs. “closed four-angled figure”).

Third, it is possible that vision and touch generate distinct representation types that, while not rationally linked, exhibit a structural correspondence that enables the subject to make reasonable inferences about whether seen and felt shapes are the same. For example, Leibniz (1765/1981, Book II, Chapter 9, Section 8) argued that in both the visual and tactual experience of a cube, eight points (the vertices) are “distinguished”. Because no such points are distinguished in visual or tactual experiences of a sphere, visual cube-experiences are structurally

---

3 See Evans (1985) for an influential defense of the rational link view based on the claim that vision and touch represent shape within a common egocentric reference frame.
more like tactual cube-experiences than tactual sphere-experiences. Still, this would not suffice for a rational link between the representations, since visual cube-experiences would exhibit this structural correspondence to tactual experiences of other properties too (e.g., oblong rectangular prisms).\(^4\)

Few philosophers have explicitly defended the type identity view. (Opponents of the view include O’Shaughnessy (1989), Martin (1992), and Prinz (2002, Chapter 5).) I’ll argue below that research on cross-modal recognition provides compelling support for this position. However, my argument will not appeal to Molyneux’s test of newly sighted subjects. A central theme in what follows is that we should not look to the newly sighted to settle whether there is an intrinsic similarity between visual and tactual shape representations in the normally sighted.

Early modern writers often treated Molyneux’s question as a thought experiment (Degenaar 1996, p. 25). However, since Cheselden’s 1728 report of the first known case of full sight restoration, many early-blind or congenitally-blind subjects have had their vision restored. In turn, it has been suggested that we might really rely on newly sighted subjects to determine whether there is an intrinsic similarity between our visual and tactual shape representations (Levin, 2008; Schwenkler, 2012, 2013; Connolly, 2013). In what follows, I argue that the challenges to using newly sighted subjects’ recognition performance for this purpose have been underestimated. Section 3 argues that for newly sighted subjects to bear evidentially on the relation between visual and tactual shape representations in normal perceivers, a condition I’ll call the *match principle* must hold. Section 4 argues that this principle is dubious.

\(^4\) More recent versions of the structural correspondence view claim that shape properties are apprehended within a “tactile field” that is distinct but analogous to the visual field (Haggard & Giovagnoli, 2011; Cheng, 2019).
3. THE MATCH PRINCIPLE

Recent empirical work has rekindled interest in Molyneux’s question. In one much-discussed study, Held et al. (2011) examined cross-modal recognition in five participants recently treated for congenital blindness. Subjects first saw or felt a sample 3D Lego shape. Next, two test shapes were presented, and subjects had to judge which of the test shapes was the same as the sample. Held et al. compared performance across three conditions: visual-visual, where sample and test shapes were all presented visually; tactual-tactual, where all were presented tactualy; and tactual-visual, where the sample shape was presented tactualy but the test shapes were presented visually.

The results were striking. Participants recognized the sample shape with high accuracy in both the visual-visual and tactual-tactual conditions (98% and 92%, respectively), but were approximately at chance in the tactual-visual condition (58%). Held et al. took these results to indicate that the answer to Molyneux’s question is “likely negative”. In a follow-up experiment, three subjects were tested again on the tactual-visual task. It was found that performance improved dramatically (to around 80% accuracy) as soon as five days after the initial test.

Does this decide Molyneux’s question in the negative? Several have expressed doubts. Schwenkler (2012, 2013) argues that for a test of newly sighted perceivers to bear on the relation between visual and tactual representations of shape, two conditions must be met. First, the subjects must be tested before they have had the chance to form associations between visual and tactual shape representations. Second, the subjects “must be able to see well enough to represent visually the shapes of the objects they are presented with” (2012, p. 186).

The problem with the Held et al. experiment, Schwenkler argues, is that we cannot be sure the second condition was met. It is unclear from subjects’ performance on the visual-visual

---

5 See Connolly (2013), Cheng (2015), Clarke (2016), and Levin (2018).
task whether they could really see well enough to represent 3D shapes in the way required for intermodal matching. For instance, subjects might have succeeded in the visual-visual condition not because they had a visual representation of the object’s global 3D shape, but because they were able to match salient local features like angles, shading patterns, or curvature segments (Schwenkler, 2013, p. 91).

Schwenkler’s two conditions are clearly necessary for an experimental test of Molyneux’s question to have evidential value. But are they sufficient? I believe they are not.

Suppose the suspicion is correct that the subjects failed to form visual representations of 3D shape. Why would this be problematic? Presumably, it is because representations of 3D shape are what underlie cross-modal shape recognition in normal perceivers. If newly sighted subjects fail to form them, then any cross-modal recognition failure on their part would not bear evidentially on the relation between visual and tactual representations in the normally sighted. More generally, the rationale behind using newly sighted subjects to test for intrinsic similarities between our visual and tactual shape representations is that the newly sighted enjoy visual and tactual representations intrinsically like ours, except they have not had the chance to learn associations between these representations. If their representations are not like ours, they cannot be relied on for this purpose.

Thus, as stated, Schwenkler’s conditions are not sufficient for Molyneux’s test to bear on the relation between visual and tactual shape representations in normal perceivers. The mere ability to visually represent the shapes to be recognized is not enough. We must also ensure that the way newly sighted subjects represent them is relevantly similar to the way normal perceivers represent them. Thus, a further condition must be met, which I will call the match principle:
MATCH PRINCIPLE: Newly sighted perceivers form visual and tactual representations of shape that are intrinsically similar to the visual and tactual representations of shape directly responsible for cross-modal shape recognition in normally sighted perceivers.

By “intrinsically similar”, I mean that the representations must closely match in their intrinsic characteristics—specifically their internal structure (e.g., imagistic or discursive format) and representational content. The match principle only covers intrinsic features because we obviously want to permit differences in extrinsic respects—namely, any cross-modal associations normally sighted perceivers have learned but newly sighted perceivers have not.6

The match principle focuses on the shape representations “directly responsible” for cross-modal recognition in normal perceivers. By this, I mean those representations that are causally relevant to cross-modal shape recognition, and not merely in virtue of causing the perceptual system to construct another representation that is more directly relevant. Suppose, for instance, that in normal perceivers, the visual system constructs two representations of shape, R1 and R2, and the formation of R1 causes it to form R2. Suppose, further, that when we need to decide whether a currently seen shape matches a previously touched shape, it is R2 (not R1) that we interrogate. Then R2 is the visual representation directly responsible for cross-modal recognition, while R1 is only indirectly responsible. The match principle requires newly sighted subjects to form representations intrinsically like R2. Why? Well, suppose that R2 bears an intrinsic similarity to tactual representations of shape in normal perceivers, and this similarity enables

---

6 Schwenkler may have intended his second condition to require not only that newly sighted subjects visually represent shape, but that they do so in the specific way that permits cross-modal comparisons in normal perceivers. If so, then we could construe the match principle as a way of elaborating Schwenkler’s second condition, rather than as an additional condition.
cross-modal recognition. If newly sighted perceivers cannot form R2, then the intrinsic similarity that enables cross-modal recognition in normal perceivers would be missing.

If the match principle is false, then the inference from newly sighted subjects’ recognition performance to intrinsic (dis)similarity between visual and tactual shape representations in the normally sighted is problematic. If newly sighted subjects fail to match seen and felt shape, then this could be because they are missing the intrinsic similarity that underlies this capacity in normally sighted subjects. And even if they succeed, it is possible that an intrinsic cross-modal similarity is present in newly sighted subjects but absent in normally sighted subjects. Perhaps normally sighted perceivers employ a more sophisticated system of shape representations in vision than in touch, and mappings between them depend on brute association. If newly sighted perceivers employ a more primitive system of visual shape representation, then cross-modal matches could be more accessible due to a closer structural correspondence between the two representational systems.

I pause to note that the importance of the match principle may vary depending on the issue we aim to resolve. If the issue is whether it is possible for someone with brand-new visual capacities to enjoy intrinsically similar shape representations across modalities—regardless of whether normally sighted perceivers do—then even if the match principle fails, successful cross-modal recognition by newly sighted subjects may still support a positive answer (although see fn. 2). If the issue is instead whether standard, normally sighted perceivers enjoy intrinsically similar shape representations across modalities, then newly sighted subjects’ recognition performance cannot resolve it unless the match principle holds. I am primarily concerned with

---

7 Failure in the task, however, arguably would not warrant a negative answer. Suppose the match principle is false because newly sighted subjects’ visual representations are different from ours. Then even if newly sighted subjects fail to match seen shape with felt, one might wonder whether someone with brand-new vision could have done so if their visual representations were intrinsically like ours. Thus, the match principle plausibly plays a key role in interpreting failures of Molyneux’s test regardless of which of the present two issues you care about.
the second issue. There is precedent for this approach, as theorists have often interpreted Molyneux’s question about newly sighted perceivers as carrying implications for the relation between visual and tactual spatial representations in the rest of us (Evans, 1985; Hopkins, 2005; Levin, 2008).

One might want to hedge the match principle in various ways. For instance, we might want to adjust it to permit newly sighted perceivers to represent shape with lower acuity than normal perceivers do. For now, however, I will not dwell on various potential weakenings of the principle. The crucial point is just that we can’t use newly sighted perceivers to test for an intrinsic similarity between visual and tactual shape representations in normal perceivers unless some form of the match principle holds. But does it? I now argue that we should have serious doubts.

4. EVALUATING THE MATCH PRINCIPLE

In this section, I first examine research on cross-modal recognition in normally sighted perceivers. This research elucidates the requirements newly sighted subjects must meet to satisfy the match principle. I then argue that it is doubtful that newly sighted subjects meet these requirements.

4.1. Cross-modal recognition
Two approaches to shape representation have dominated psychological research on object recognition: *structural* models and *holistic* models. I briefly review these before turning to evidence on cross-modal shape recognition.

Structural models hold that the representation of an object’s shape specifies a particular decomposition of the object into parts, the intrinsic shapes of the parts, and the spatial relations the parts stand in to one another (Biederman, 1987; Hummel, 2003, 2013). In structural models these aspects of shape are represented independently: The representation of an object’s global shape contains discrete constituents that represent the intrinsic shapes of its parts and spatial relations between parts. Accordingly, structural descriptions make explicit geometrical similarities between objects that either share intrinsic part shapes but differ in spatial configuration, or that differ in intrinsic part shapes but have a similar spatial configuration (Barenholtz & Tarr, 2008; Green, 2019) Although it is not obligatory, structural models tend to represent shape in object-centered reference frames. The intrinsic shapes of parts are encoded in reference frames centered on the intrinsic axes of the parts (Feldman & Singh, 2006; Lowet et al., 2018), while spatial relations among parts are specified in terms of their relations to one another, rather than their relations to the perceiver’s viewpoint. Accordingly, structural representations are predicted to remain relatively invariant to certain changes in perspective on an object.

Holistic models differ from structural models in that they do not prioritize any specific decomposition of an object into mid-sized parts. Instead, the representation of shape specifies the numerical coordinates of local “critical features”, such as vertices, curvature extrema, or inflection points (e.g., Poggio & Edelman, 1990; Ullman & Basri, 1991; Riesenhuber & Poggio, 2002). No particular part decomposition is prioritized, since any way of dividing an object into
parts is compatible with the same collection of vertices and curvature maxima. Holistic models are also viewer-centered: Critical features are located in a reference frame centered on the perceiver. Thus, the representation of shape changes significantly with shifts in perspective on an object.

While structural and holistic representation schemes may represent the same shape properties (or approximately so), they represent them by way of different descriptive contents. Structural schemes represent an object’s global shape by describing the intrinsic shapes of its parts along with their spatial relations, both in object-centered reference frames. Holistic schemes represent an object’s global shape by describing the numerical coordinates of its vertices and other critical features in a viewer-centered reference frame. Another difference is that structural models typically represent part shapes by means of shape primitives applying to larger regions of the shape (e.g., straight vs. curved axis; parallel vs. tapered sides), while holistic models take local contour features as primitive (e.g., vertex at position \((x, y, z)\)). So even if the two schemes represent the same shape property, they differ in the description under which they represent it. Thus, structural and holistic representations differ intrinsically in the sense relevant to evaluating the match principle.

The simple question of whether shape representation is structural or holistic is probably misconceived. Both approaches have empirical support. Either kind of representation might be used depending on the level of perceptual processing, the type of shape to be represented, or the conditions of attention (Hummel, 2013; Gauthier & Tarr, 2016). Still, the balance of evidence suggests a progression from more holistic representations to more structural representations during sensory processing (Hummel, 2003; Gauthier & Tarr, 2016, p. 379). Structural
representations are likely to reside in late processing areas, which show greater invariance to position, scale, and orientation (Kobatake & Tanaka, 1994).

Our interests are more specific than the general issue of how perception represents shape. We are concerned with the representations that underlie cross-modal shape recognition in normally sighted perceivers. What sorts of representations explain our ability to visually identify a shape previously encountered only through touch? In this area, recent research reveals a central role for structural, view-invariant representations.

I start with the evidence for view-invariance. Lacey et al. (2007) had participants study four objects like the one in Figure 1. The study phase was either solely visual or solely tactual, and only one “view” was permitted. Subjects either saw the objects from a fixed viewpoint, or haptically explored them at a fixed orientation. After the study phase, subjects encountered another object and were asked which of the four learned objects it was. The test object could either be presented to the same modality as the study phase or to the other modality, and it could either be presented in the same orientation or could be rotated 180°. For the within-modal cases, recognition was view-dependent. Participants were more accurate in visually (tactually) recognizing an object that they had seen (felt) at the same orientation during study. However, cross-modal recognition was view-independent. When study and test modalities differed, recognition was no less accurate at the rotated orientation than at the original orientation.
The view-independence of cross-modal shape recognition has been reproduced in other studies (Lacey et al., 2010; Ueda & Saiki, 2012). This suggests that the representations responsible for transfer of recognition from vision to touch, and vice versa, are probably view-invariant rather than view-dependent. They support the same accuracy regardless of orientation.

One way to explain these results is via a progression from view-dependent to view-invariant representations during perceptual processing (Lacey & Sathian, 2014). The idea is that cross-modal tasks only recruit later, view-invariant representations of shape, while within-modal tasks can recruit earlier, view-dependent representations. The latter representations facilitate better performance when the object is presented at familiar orientations. Lacey and Sathian (2014) identify the lateral occipital complex (LOC) as the most likely candidate for housing view-invariant representations. LOC is relatively late in visual processing, and there is evidence that it can exhibit view-invariant responses to shape (James et al., 2002). It is also active during both visual and tactual perception of shape, but not when touching simple textures or seeing

**Figure 1.** A sample object from the Lacey et al. (2007) study. The same shape is shown at four orientations. Source: Lacey et al. (2007). Reproduced under the terms of the Creative Commons Attribution License.
scrambled noise (Amedi et al., 2001), making it a natural place to expect the kind of coordination involved in cross-modal shape recognition. Finally, there is evidence that lesions to LOC produce deficits in both visual and haptic object recognition (James et al., 2006).

Recent work has examined the nature of visual and haptic shape representations in LOC. This research provides compelling support for part-based representation, the other main aspect of a structural scheme (alongside view-invariance). Erdogan et al. (2016) had participants either view or haptically explore objects while fMRI recordings were taken of LOC and other sensory areas. As expected, both visual and haptic shape perception activated LOC. To assess whether the LOC’s representations were part-based or holistic, Erdogan et al. examined responses to a collection of 16 multi-part objects. While gross spatial configuration of parts was held fixed, each individual part could adopt one of two intrinsic shapes (see Figure 2). Thus, every object had a unique global shape, but individual part shapes were shared among eight objects. Erdogan et al. reasoned that if LOC explicitly represents intrinsic part shapes, then these representations should be shared between objects that share a part, even if their global shape differs. To test this, they asked whether it was possible for a linear classifier to predict the shape of the object that produced a given LOC response on the basis of LOC responses to the other objects in the stimulus set, which differed in global shape but shared one or more individual part shapes. They found that this could be done, regardless of whether the responses were produced visually or haptically.8 A natural explanation is that activity patterns in LOC explicitly represented the part shapes of the other objects, and these patterns were partially reproduced for the object being decoded.

---

8 An earlier study using a similar method also found support for part-based representations in LOC in the case of vision, but did not examine haptic perception (Guggenmos et al., 2015).
The hypothesis that shape is represented in a part-based manner in both vision and touch is also supported by an earlier behavioral study from the same group. Using the same collection of shapes (Figure 2), Erdogan et al. (2015) found that participants’ ratings of shape resemblance were reliably determined by similarity in part shapes, and ratings between two shapes were consistent regardless of whether the shapes were both seen, both touched, or one touched and one seen.

It might be questioned whether LOC activity during haptic shape perception merely reflects visual imagery, not haptic processing per se. However, this is unlikely. Amedi et al. (2001) compared LOC activity during haptic perception and visual imagery and found far weaker activation in the latter case. Furthermore, Amedi et al. (2010) found similar magnitudes of LOC activity during haptic shape perception for sighted and congenitally blind subjects.

Figure 2. Stimuli used in the Erdogan et al. (2016) study. Reprinted from Erdogan et al. (2016) with permission from MIT Press Journals.
Thus, LOC represents shape in both sight and touch. Moreover, it probably employs a part-based scheme in both cases. Here, then, is a hypothesis concerning the processes responsible for cross-modal shape recognition. The visual and haptic systems initially generate holistic, view-dependent representations of shape. These can be used for intra-modal shape recognition, and they enable better recognition at familiar viewpoints. At later stages, both modalities produce part-based, view-invariant representations. Likely implemented by LOC, these representations account for both cross-modal recognition and intra-modal recognition at unfamiliar orientations. These structural representations are directly responsible for cross-modal recognition. Earlier view-dependent representations are indirectly responsible in virtue of their role in producing structural representations.

4.2. Impairments in the newly sighted

The match principle requires that newly sighted subjects form shape representations intrinsically like the ones directly responsible for cross-modal recognition in normal perceivers. Section 4.1 adduced evidence that the representations that fill this role for normal perceivers are view-invariant and part-based. Thus, for the match principle to hold, newly sighted perceivers must be capable of forming representations of this type. Unfortunately, the available evidence on newly sighted subjects casts doubt on their ability to do this.

Newly sighted subjects exhibit general impairments of mid-level and high-level visual processing. While they can discriminate local contours, motion, and simple 2D forms, they exhibit marked deficits in contour integration, perception of 3D shape, and object recognition. In some cases, these impairments persist for years following sight restoration (Huber et al., 2015).
McKyon et al. (2015) studied impairments of mid-level visual processing in the newly sighted. The participants (11 congenitally blind subjects who had sight surgically restored) were shown a display of objects and had to identify the item that differed from the rest. The oddball could differ either in a feature designed to tap early vision: for example, color, size, or 2D shape; or in one designed to tap mid-level vision: for example, amodally completed 2D shape, orientation of a 3D volume, or 2D shape defined by subjective contours (see Figure 3). Performance was compared to that of normally sighted control subjects who saw blurred versions of the same stimuli to compensate for newly sighted subjects’ reduced acuity. It was found that newly sighted subjects performed similarly to control subjects on low-level tasks, but were indistinguishable from chance, and significantly less accurate than controls, on three of the four mid-level tasks. The lone exception was the shape-from-shading condition, although in this case it was found that newly sighted subjects’ reaction time increased with display size while normal subjects’ did not, suggesting the use of a different strategy.

![Figure 3](image-url) Oddball detection tasks studied by McKyon et al. (2015). Reprinted from McKyon et al. (2015) with permission from Elsevier.

This impairment of mid-level visual capacities should make us skeptical of newly sighted subjects’ ability to form structural, view-invariant shape representations in vision, since these
representations are thought to be formed in late stages of processing. However, further evidence gets at this issue more directly.

First, there is evidence that newly sighted subjects struggle to recognize shapes across changes in 3D orientation. Huber et al. (2015) report various tests of subject MM, who lost sight at age 3.5 and had sight restoration surgery at age 43. In one test, MM was shown images of two 3D objects on opposite sides of the screen and was asked to indicate whether the two images depicted versions of the same shape at different orientations in depth. MM’s performance in this task was indistinguishable from chance and significantly worse than control subjects, suggesting an impaired ability to form or access view-invariant representations. MM was also significantly worse than controls—though better than chance—in matching 2D shapes at different orientations in the picture plane.9

Another study found that two subjects who had their sight restored following years of blindness struggled to recognize familiar objects from atypical viewpoints. Šikl et al. (2013) studied both subject MM and another subject, KP, who went blind at age 17 and had vision surgically restored 53 years later. Both KP and MM had trouble recognizing common items like bicycles or coffee mugs from atypical perspectives. KP accurately identified 36/39 objects from canonical viewpoints, but only 6/10 from atypical viewpoints (MM recognized 4/10). KP’s decades of blindness may have impacted his ability to form the view-invariant representations needed to recognize objects from unfamiliar viewpoints.

So there is evidence that newly sighted subjects struggle to visually form view-invariant shape representations, if they can form them at all. This puts pressure on the idea that the sorts of

---

9 In a study of another patient, SRD, Ostrovsky et al. (2006) reported successful 3D shape matching across changes in depth. However, these tests occurred twenty years after SRD’s sight restoration surgery. Moreover, Ostrovsky et al. note that SRD took longer than normal subjects to make matches, suggesting the use of a different strategy (2006, pp. 1012–1013).
visual shape representations directly responsible for cross-modal recognition in normal perceivers—which we have seen are view-invariant—are available to the newly sighted. There is also reason to think that newly sighted subjects struggle with the other key ability involved in structural shape representation: part segmentation.

Structural representation demands sophisticated visual parsing abilities. Specifically, it requires forming *hierarchical* representations of objects, where perceptual organization is specified at multiple levels (Palmer, 1977). At one level, a coffee mug is represented as a single, unified individual, while at a subordinate level the handle is represented as distinct from the cylindrical portion. Parsing at the subordinate level also requires sensitivity to sophisticated principles of *segmentation*. Boundaries between parts are not always marked by the luminance or color discontinuities characteristic of physical edges between disconnected surfaces. Instead, they are often signaled by geometrical cues. Boundaries are found at concave curvature minima, and part segmentation involves linking these minima along the shortest paths possible (Singh & Hoffman, 2001). Furthermore, when parts are separated by shading or color discontinuities, this cannot be taken to preclude a higher level of organization where the separated regions are integrated into a common whole.

To my knowledge, no study has directly tested the perception of parthood in newly sighted subjects. However, the available data on parsing in newly sighted subjects indicates that they struggle with the kinds of abilities needed for part decomposition. Ostrovsky et al. (2009) examined visual parsing in three subjects within three months following sight restoration. They were shown various displays and asked to report how many objects the display contained. This ability was greatly impaired. For a 3D cube with differently shaded faces, subjects judged that each face corresponded to a separate object. In the case of partially overlapping 2D shapes, all
closed loops were treated as separate objects. Ostrovsky et al. found that subjects’ performance closely matched a simple algorithm that segmented regions according to shared luminance or hue.\(^\text{10}\)

This simple parsing rule would be inadequate to the kind of perceptual organization involved in part-based shape representation. The problems are twofold. First, recall that part boundaries are not always marked by luminance or color discontinuities. The simple parsing rule would be blind to such boundaries. Second, when parts are separated by luminance discontinuities (e.g., because of differential shading or color), normal perceivers can still integrate them into a larger whole. Newly sighted subjects appear unable to do this, indicated by their failure to treat the facets of a 3D volume as parts of the same object. Thus, the parsing abilities of newly sighted subjects are poorly suited to structural shape representation.

These results cast doubt on the idea that newly sighted subjects can form structural representations of shape within vision. This is a problem for the match principle, since these are the sorts of representations that we know are responsible for cross-modal shape recognition in normal perceivers.

The data on newly sighted subjects remains scant, and there are individual differences (Ostrovsky et al., 2006; Huber et al., 2015). The crucial point is just that we should have a healthy skepticism about newly sighted subjects’ ability to form the relatively sophisticated representations responsible for cross-modal recognition in normal perceivers. It is a live possibility that newly sighted subjects are limited to the sorts of holistic, view-dependent representations formed at earlier stages of visual processing, and this explains their deficits in both reidentifying a shape across viewpoints and sophisticated visual parsing. If so, cross-modal

---

\(^\text{10}\) Note that while subjects’ parsing abilities improved when dynamic motion cues were added, these cues would be ineffective for parsing an object into parts, since the parts of an object typically move together.
recognition failures in the newly sighted should could not count against an intrinsic similarity between visual and tactual shape representations in the rest of us. Perhaps there is an intrinsic similarity, but it arises at the level of structural representations to which newly sighted subjects lack access.

5. CAN THE TEST BE FIXED?

I am not the first to observe that visual deficits in newly sighted subjects pose a problem in interpreting empirical tests of Molyneux’s question. However, some have suggested that such experiments may still have evidential import if we appropriately refine Held et al.’s (2011) methodology. I now argue that if newly sighted subjects are unable to form part-based, view-invariant shape representations through vision, then the proposed refinements are inadequate.

Schwenkler (2012, 2013) argues that the primary problem with the Held et al. experiment is that the stimuli were too complex and viewing opportunities too restricted. Recall that the stimuli were 3D Lego blocks, and subjects saw the blocks from a single viewpoint. Schwenkler argues that newly sighted subjects may have been unable to visually represent the shapes of these blocks, given that the shapes were complex and the depth cues available from a fixed viewpoint are limited. He suggests refining the experiment in either of two ways. First, we might stick with 3D shapes, but present them rotating on a platform (or while subjects are free to move around them). Second, we might test whether cross-modal transfer is possible for 2D raised-line drawings, since these would plausibly place fewer visual demands on newly sighted subjects.
Both of these proposals have faced pushback. Connolly (2013) doubts the feasibility of the first option, since he thinks it is questionable whether newly sighted subjects would be any more sensitive to dynamic depth cues than they are to static cues. While Connolly favors the second option, Cheng (2015; see also Schwenkler, 2019) argues against it on the grounds that the haptic system has difficulty processing the shapes of raised-line drawings.

I believe a common concern saddles both suggested refinements. It is that they are tailored to allow newly sighted subjects to visually represent, *in some way or other*, the shapes of the objects to be recognized, but they cannot ensure that newly sighed subjects represent these shapes in the *specific way* responsible for cross-modal recognition in normally sighted perceivers.\(^\text{11}\)

Suppose that newly sighted subjects are visually limited to holistic, view-dependent representations. Then it is plausible that they would find it easier to represent 2D shapes and 3D shapes seen from multiple viewpoints. Regarding the former, this is because the object’s complete shape is accessible from a single perspective. Regarding the latter, this is because multiple views may allow their visual systems to form a series of view-dependent representations of the object, which specify aspects of shape that were not represented in the initial viewpoint. However, in neither case would this ensure that the subjects formed a structural, view-invariant representation.

First consider the 2D case. 2D shape can be represented either holistically (as a vector of local feature coordinates) or structurally (as a collection of parts described by higher-level shape primitives). Since we know that normal perceivers rely on structural representations for cross-

\(^{11}\) Millar (2019) has suggested that newly sighted subjects might have basic shape phenomenology but fail to represent external shape properties. Although nothing I have said conflicts with this view, the present point is different. Even if newly sighted subjects *do* represent shape, there is no guarantee that they do so in the way responsible for cross-modal recognition in normal perceivers.
modal recognition of 3D shapes, there is no reason to think that they would not also rely on them for 2D shape. Thus, if newly sighed perceivers lack access to structural representations, then testing them on 2D shapes is insufficient. While they might find it easier to represent 2D shapes, this wouldn’t ensure that they represented them in the specific way responsible for cross-modal recognition in normal perceivers.

Similar points apply to 3D shapes seen from multiple viewpoints. Newly sighted subjects might find it easier to represent and recognize a 3D shape if it is seen from multiple viewpoints because they would have the chance to form a series of view-dependent representations. However, this would not ensure that they represented the shape by means of a structural, view-invariant (i.e., object-centered or part-centered) representation.

Interestingly, simplifying the stimuli or enriching the viewing conditions may actually be counterproductive. By testing newly sighted subjects on stimuli that are easier to represent and reidentify in a holistic scheme, we might fail to notice impairments in the more sophisticated schemes of shape representation responsible for cross-modal recognition in normal perceivers.

To sum up: Cross-modal recognition in normal perceivers seems to depend on structural, view-invariant shape representations. Because it is doubtful that newly sighed subjects form these representations, the match principle is dubious. If the match principle is false, then simplifying the stimuli or enriching the viewing conditions is insufficient to restore the evidential import of experimental tests of Molyneux’s question.

---

12 Thus, the fact that subject MM in Fine et al. (2003) could recognize a cube when the figure was set in motion does not show that the property was represented in a structural, view-invariant manner. It is equally possible that recognition was based on the sort of matching to view-dependent representations involved in holistic models of recognition.
6. A CASE FOR TYPE IDENTITY

6.1. Support for type identity

An upshot of the foregoing discussion is that newly sighted subjects are a weak source of evidence concerning the relation between visual and tactual shape representations in normally sighted perceivers. However, there are other sources of evidence at our disposal. I now argue that a strong case can be made for the view that the visual and tactual representations of shape directly responsible for cross-modal recognition in normal perceivers are intrinsically similar. In fact, I suggest they are type-identical.

My argument is as follows: The evidence indicates that the visual and tactual representations directly responsible for cross-modal recognition in normal perceivers exhibit both physiological overlap and shared functional role. The best explanation of these facts, I contend, is that they are type-identical.

We have seen that the lateral occipital complex (LOC) responds both to seen and touched shape. However, this falls short of establishing that LOC responds in the same way (i.e., with similar activity patterns) in both cases. Fortunately, another experiment from the Erdogan et al. (2016) study mentioned above provides evidence for the stronger thesis. Erdogan compared the patterns of BOLD activity in LOC when a complex shape was touched versus seen. To gauge whether the response patterns were similar in both cases, they asked whether the activity produced while seeing a shape was better correlated with that produced while touching that same shape than with that produced while touching another shape in the stimulus set. In right LOC, the former correlations were significantly higher for all 12 subjects, while in left LOC this was true for 10/12 subjects. These results suggest that not only does LOC happen to process both visual
and haptic information about shape—it responds in a similar way regardless of whether a given shape property is perceived visually or tactually.

Further evidence for physiological overlap derives from neural adaptation effects. When a stimulus is repeated, neural regions responsive to the stimulus exhibit reduced activation—a phenomenon known as “repetition suppression”. Tal and Amedi (2009) exploited this to test for physiological overlap in neural responses to visually and haptically perceived shape. Subjects saw a picture of an object, then haptically explored either a same-shaped object or a different-shaped object. Again, consistent with physiological overlap, fMRI recordings revealed significantly greater response reduction when the same object was repeated, and this cross-modal adaptation effect was strongest in left LOC.

While the issue is by no means decisively resolved, vision and touch appear to produce shape representations with substantial physiological overlap in LOC. However, this still does not show that the representation types produced in the two cases are identical. After all, it is possible that LOC exhibits similar responses to a shape regardless of whether it is presented visually or haptically, but that this neural state implements different representation types in the two cases (Levin, 2018, p. 600), just as bits of computer hardware can be used to implement different representations depending on the program being run. To determine whether the representation types are the same, we must also determine whether they play the same role in the perceiver’s psychology. Behavioral evidence suggests that this is so.

Cross-modal shape recognition is view-independent even when subjects are familiarized with only one perspective on an object, suggesting that it relies on view-invariant representations. However, this is consistent with two possibilities: (1) vision and touch produce tokens of distinct view-invariant representation types that are compared with one another during
cross-modal recognition. When the two representations are deemed to match, the subject identifies a seen shape as the same as one she touched, or vice versa; (2) vision and touch produce tokens of the same view-invariant representation type. Because the same representation is produced in both cases, the subject identifies a seen shape as the same as one she touched, and vice versa. A subsequent study by Lacey et al. (2009) provides support for the latter option. However, some background is needed before considering its findings.

While within-modal shape recognition of novel objects often starts out view-dependent, it can become less view-dependent with practice. Costs in accuracy or reaction time for recognizing an object across changes in orientation can be reduced or eliminated following a training period where one perceives the object from different views. There are many reasons this might occur. Additional exploration of an object might increase the detail or accuracy of its view-invariant representation. Practice with multiple views might also speed the perceptual formation of these view-invariant representations (e.g., Bar, 2001), speeding reaction time. Importantly, regardless of which specific story is correct, these training-induced effects on view-independence are indicative of the functional role of view-invariant representations. They reveal how these representations are updated or refined given new perceptual information, and/or how quickly they are formed given sensory input.

If our visual and tactual view-invariant representations are type-identical, then updating one’s visual view-invariant representation of a given shape property (e.g., its detail, precision, or speed of formation) must suffice for the same updates to one’s tactual representation of that property. Lacey et al. (2009) showed just this type of carryover effect in the case of learned view-independence. The experiment had three stages. First, view-dependence was tested within vision and touch by examining subjects’ accuracy in recognizing shapes either at the same
orientation or in a new orientation. In both modalities, participants were less accurate when orientation changed between initial exposure and test. Next, subjects completed a training phase where they encountered the shapes at different orientations. For half of the subjects, this phase involved only visual presentations, while the others encountered only tactual presentations. Finally, the effect of training was assessed by again testing for view-dependence in both visual and haptic recognition. The critical finding was that both visual and haptic recognition became view-independent following training, regardless of the modality in which training took place. Visual training produced haptic view-independence, and vice-versa (see Figure 4). This suggests that the updates to view-invariant representations involved in learned view-independence immediately carry over across modalities. At least in this respect, visual and tactual view-invariant representations seem to have a shared functional role: Revise one of them, and you revise the other.

![Figure 4](image-url)  
Figure 4. Results of the Lacey et al. (2009) experiment. Final recognition performance is independent of orientation, regardless of whether the learning phase was visual or haptic. Reprinted from Lacey et al. (2009) with permission from Springer Nature.

I argued in Section 4.1 that both vision and touch produce structural, view-invariant representations in late stages of processing. This section has adduced evidence that these representations physiologically overlap and have a shared functional role. Together, these strands
of evidence suggest a straightforward account: The visual and tactual shape representations directly responsible for cross-modal recognition in normal perceivers are type-identical. We visually recognize a previously touched object, and vice versa, because the very same representation types are produced in both cases. *A fortiori*, our visual and tactual shape representations exhibit intrinsic similarity. This is so regardless of whether that similarity is replicated in newly sighted subjects.

One might wonder whether the ability to form type-identical shape representations across modalities is present at birth or instead results from a history of cross-modal coordination or acquired habit. One reason to think that the ability may be innate is that reliable cross-modal shape recognition is possible within mere hours of birth (Streri & Gentaz, 2003). Nevertheless, more data is needed to reveal whether infants’ recognitional capacities draw on the sorts of structural, view-invariant shape representations emphasized here.

### 6.2. Objections and replies

I consider two sorts of objections to the type identity view. Those of the first sort challenge my appeal to evidence of physiological overlap. The second sort of objection challenges type identity by highlighting modal differences in shape phenomenology.

One might object that the presence of neurons that respond to both visual and tactual input is not really indicative of shared representations across modalities. One might insist that cross-modally activated neural states actually implement sense-specific representations. Their apparent multisensory nature is due to causal relations enabling a sense-specific representation in modality 1 to activate a separate, sense-specific representation in modality 2 (Prinz, 2002, p.
136). If so, the representations in LOC would not be multisensory, but rather vision-specific representations activated via connections from haptic brain areas such as somatosensory cortex.

The challenge for this view is to explain what makes the representations in LOC vision-specific, given that they are produced in the absence of retinal input (Amedi et al., 2001), and encode haptically perceived shape even for congenitally blind subjects (Amedi et al., 2010). Moreover, the LOC plays a key role in both visual and tactual object recognition (James et al., 2006), suggesting that it is computationally integrated with processing in both modalities.

Another objection in a similar vein would allege that the states of LOC do not really implement perceptual representations of shape, but rather recognitional concepts of shape. It is unsurprising that we can apply the same recognitional concept (e.g., SPHERE) on the basis of either vision or touch. If LOC were simply the neural basis of these concepts, then it would be equally unsurprising to see it activated both visually and tactualy.

However, this is not a plausible interpretation of the evidence. Erdogan et al. (2016) tested subjects on novel shapes for which they would have lacked recognitional concepts. So if LOC implemented shape representations in this study (and, recall, it was possible to decode object shape from LOC activity patterns), these were perceptual representations, not recognitional concepts.

I turn to the second sort of objection to the type identity view. I have suggested that visual and tactual shape perception employ some of the same representations. One might argue, however, that this fits poorly with spatial phenomenology, which has been argued to differ starkly between vision and touch (Block, 1996; O’Dea, 2006).

An initial reply is to point out that I have not committed to any view of the relation between the phenomenology of shape and the perceptual representation of shape. It is possible,
for all I have said, that although vision and touch share the same shape representations, the phenomenology associated with these representations is different in the two cases. However, I think we can say more than this. I contend that the view defended here fits the phenomenological data rather well.

Note, first, that if shape phenomenology were the same across modalities, we might overlook this due to the distracting fact that shapes are always co-presented with colors in vision, which have no counterpart in touch. Thomas Reid flagged a similar issue when considering whether a blind man might be able to conceive of visible figure:

[The] blind man’s notion of visible figure will not be associated with colour, of which he hath no conception; but it will perhaps be associated with hardness or smoothness, with which he is acquainted by touch. These different associations are apt to impose upon us, and to make things seem different, which in reality are the same. (Reid, 1764/1997, Chapter 6, Section 7).

However, one might object that the apparent differences in visual and tactual shape phenomenology could not be solely due to differences in other features with which shapes are associated. It is implausible that whenever we are impressed by modal differences in shape phenomenology, this is just because we are unable to introspectively disentangle shape from color.

Fortunately, the present view has resources to accommodate modal differences in shape phenomenology. On the proposal discussed in Section 4.1, there are multiple levels of shape representation in sensory processing. There is a holistic, view-dependent stage, followed by a structural, view-invariant stage. Because the view-dependent stage is modality-specific,
differences at this level may ground modal differences in shape phenomenology. These modality-specific representations almost certainly encode shape in different reference frames (eye-centered or head-centered in the case of vision, hand-centered in the case of touch). They may also describe shape using different geometrical primitives.\(^\text{13}\)

One option would be to hold that shape phenomenology is *wholly* grounded in these modality-specific, view-dependent representations. However, if this were the claim, then one might reasonably object that the present view is silent on the issue that really concerned Locke, Berkeley, et al.—namely, whether there is a phenomenal similarity between the representation of shape in sight and touch.

So that is not the position I favor. It seems to me that while our overall experiences of shape differ between sight and touch, they also have aspects in common, and shared structural representations may explain the commonalities. What it is like to visually experience an object as composed of a curved handle side-attached to a cylinder may well be the same as what it is like to tactually experience something as having this property. We often overlook this shared aspect not merely because of Reid’s point that there are modal differences in the other properties we are aware of, but also because we experience shape in multiple ways within a modality. In some respects, our experience of shape is phenomenally alike between vision and touch, while in other respects it is not. Thus, I conjecture that the visual and tactual shape representations directly responsible for cross-modal recognition are type-identical, and that this grounds a common aspect of shape phenomenology between vision and touch.

**ACKNOWLEDGEMENTS**

\(^\text{13}\) This view permits the possibility of rational or structural correspondences between modality-specific representations at earlier stages. Thus, it may be possible to integrate the type identity view with recent structural correspondence theories (e.g., Haggard & Giovagnoli, 2011) if the two views characterize different levels of shape processing.
Thanks to two anonymous reviewers for their insightful comments. I have benefited from discussing these issues with Tyler Brooke-Wilson and Alex Byrne. Thanks also to audiences at Brown University and the MIT work-in-progress seminar, where portions of this material were presented.

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., 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.

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

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

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

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

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

Campbell, J. (1996). Molyneux’s question. *Philosophical Issues, 7*, 301-318.

Cheng, T. (2015). Obstacles to testing Molyneux’s question empirically. *i-Perception, 6*(4), 1-5.

Cheng, T. (2019). On the very idea of a tactile field, or: A plea for skin space. In T. Cheng, O. Deroy, & C. Spence (eds.), *Spatial Senses: Philosophy of Perception in an Age of Science*. London: Routledge.

Clarke, S. (2015). Investigating what felt shapes look like. *i-Perception, 7*(1), 1-6.

Connolly, K. (2013). How to test Molyneux's question empirically. *i-Perception, 4*(8), 508-510.

Degenaar, M. (1996). *Molyneux’s Problem: Three Centuries of Discussion on the Perception of Forms*. Dordrecht: Kluwer.

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.

Gauthier, I., & Tarr, M. J. (2016). Visual object recognition: Do we (finally) know more now than we did? *Annual Review of Vision Science, 2*, 377-396.

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

Guggenmos, M., Thoma, V., Cichy, R. M., Haynes, J. D., Sterzer, P., & Richardson-Klavehn, A. (2015). Non-holistic coding of objects in lateral occipital complex with and without
Haggard, P., & Giovagnoli, G. (2011). Spatial patterns in tactile perception: is there a tactile field? *Acta Psychologica, 137*(1), 65-75.

Held, R., Ostrovsky, Y., de Gelder, B., Gandhi, T., Ganesh, S., Mathur, U., & Sinha, P. (2011). The newly sighted fail to match seen with felt. *Nature Neuroscience, 14*(5), 551-553.

Hopkins, R. (2005). Molyneux’s question. *Canadian Journal of Philosophy, 35*(3), 441-464.

Huber, E., Webster, J. M., Brewer, A. A., MacLeod, D. I., Wandell, B. A., Boynton, G. M., Wade, A. R., & Fine, I. (2015). A lack of experience-dependent plasticity after more than a decade of recovered sight. *Psychological Science, 26*(4), 393-401.

Hummel, J. E. (2003). The complementary properties of holistic and analytic representations of object shape. In G. Rhodes & M. Peterson (eds.), *Perception of Faces, Objects, and Scenes: Analytic and Holistic Processes*. Westport, CT: Greenwood, pp. 212-234.

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

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. Ballesters (eds.), *Touch and Blindness: Psychology and Neuroscience*. Mahwah, NJ: Lawrence Erlbaum, pp. 139-155.

Kobatake, E., & Tanaka, K. (1994). Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *Journal of Neurophysiology, 71*(3), 856-867.

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.

Leibniz, G. (1765/1981). *New Essays on Human Understanding*. P. R. Remnant & J. Bennett (eds.). Cambridge: Cambridge University Press.

Levin, J. (2008). Molyneux’s question and the individuation of perceptual concepts. *Philosophical Studies, 139*(1), 1-28.

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

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

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

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

Matthen, M., & Cohen, J. (2020). Many Molyneux Questions. *Australasian Journal of...*
McKyon, A., Ben-Zion, I., Doron, R., & Zohary, E. (2015). The limits of shape recognition following late emergence from blindness. Current Biology, 25(18), 2373-2378.

Millar, B. (2019). Learning to see. Mind & Language. DOI: 10.1111/mila.12263.

Norman, J. F., Norman, H. F., Clayton, A. M., Lianeekhammy, J., & Zielke, G. (2004). The visual and haptic perception of natural object shape. Perception & Psychophysics, 66(2), 342-351.

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.

Ostrovsky, Y., Andalman, A., & Sinha, P. (2006). Vision following extended congenital blindness. Psychological Science, 17(12), 1009-1014.

Ostrovsky, Y., Meyers, E., Ganesh, S., Mathur, U., & Sinha, P. (2009). Visual parsing after recovery from blindness. Psychological Science, 20(12), 1484-1491.

Palmer, S.E. (1977). Hierarchical structure in perceptual representation. Cognitive Psychology, 9, 441-474.

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.

Reid, T. (1764/1997). An Inquiry into the Human Mind. Derek R. Brookes (ed.). Edinburgh: Edinburgh University Press.

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

Schwenkler, J. (2012). On the matching of seen and felt shape by newly sighted subjects. i-Perception, 3(3), 186-188.

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.

Šikl, R., Šimecek, M., Porubanova-Norquist, M., Bezdicek, O., Kremlácek, J., Stodulka, P., Fine, I., & Ostrovsky, Y. (2013). Vision after 53 years of blindness. i-Perception, 4(8), 498-507.

Singh, M., & Hoffman, D.D. (2001). Part-based representations of visual shape and implications for visual cognition. Advances in Psychology, 130, 401-459.

Streri, A., & Gentaz, E. (2003). Cross-modal recognition of shape from hand to eyes in human newborns. Somatosensory & Motor research, 20(1), 13-18.

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.

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.

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