




Structural correspondence in Molyneux's subjects

Tony Cheng^a  (h.cheng.12@ucl.ac.uk)

Abstract

The historical Molyneux's question – roughly, whether congenital blind subjects can visually identify shapes in front of them right after being made to see – is having its renaissance in recent years (Ferretti & Glenney, 2021). While there have been many different formulations of it, and many attempted answers as well, no clear consensus has been reached. Moreover, although arguably both memory and imagination are involved in the process, their roles in the Molyneux's task have not been adequately discussed. In this paper, I focus on a specific version of Molyneux's question, propose a specific hypothesis in relation to that version, and sketch the roles of structural correspondence, memory, and imagination in the relevant hypothesis. One key moral of this discussion is that while Molyneux's question has primarily been regarded as a perceptual puzzle, other kinds of mental episodes such as memory and imagination have to be taken into account as well in order to have a more satisfying answer to versions of Molyneux's question (Macpherson & Dorsch, 2018).

Keywords

Imagination · Memory · Molyneux's question · Multisensory knowing-how · Structural correspondence

This article is part of a special issue on “Molyneux's question today”, edited by Gabriele Ferretti and Brian Glenney.

1 Many Molyneux's questions

In the western history of philosophy, there are many famous questions, problems, or puzzles. The so-called “Molyneux's question” (or sometimes “Molyneux's problem”) is one amongst them. Before imposing any interpretation, a direct quote of the origin is often the best starting point:

Suppose a Man born blind, and now adult, and taught by his touch to distinguish between a Cube, and a Sphere of the same metal, and

^a Waseda University.

nighly of the same bigness, so as to tell, when he felt one and t'other, which is the Cube, which the Sphere. Suppose then the Cube and Sphere placed on a Table, and the Blind Man to be made to see. Quære, Whether by his sight, before he touched them, he could now distinguish, and tell, which is the Globe, which the Cube. (Locke, 1975)

This began as personal correspondences between William Molyneux and John Locke in 1688, and it appeared officially in the second edition of Locke's *An Essay Concerning Human Understanding*, as quoted above.

Although the passage is short, it has incurred many discussions since then, and been interpreted as about different issues, such as the epistemology of concepts, the relation between concepts and perceptual experiences, and amodal spatial representations, to name a few. Many famous figures in the western history of philosophy, including Leibniz, Berkeley, Reid, Diderot, Condillac, Lotze, etc., have taken their stands on this. Here I shall not venture to survey this broad interpretative territory.¹ The specific version of Molyneux's question I will focus on is called the "idealist version," according to which the subjects are made to see *right away by perfect surgeries or procedures*, which guarantees that their visual capacities are as good as they can get as long as they are within the standard range of human vision.² This thought experiment version implies that no perceptual learning is needed in order to gain such capacities (cf. Connolly, 2019).

Given that so many different issues have been discussed in the relevant literature, the aim of this paper needs to be stated clearly at this early stage. The major research questions are these: do vision and touch generate the same or distinct representation type(s) concerning shapes? If distinct, how are they related? Below I will argue for "distinctness," and provide a positive account of how the relevant visual and tactile representations relate. In due course, I will need to argue against other views, especially E. J. Green's case for "sameness."

Now with this framework and aim, this paper will go as follow. Section 2 will discuss various versions of intrinsic similarity (Green, 2022), and will critically evaluate Green's case for his preferred version (Type identity). Section 3 will develop and defend my preferred version (Structural correspondence), and discuss how memory and imagination are involved in this version of Molyneux's question. Section 4 concludes with the notion of "multisensory knowing-how" and gesture a future direction of research.

¹ For interpretations and expansions, for example see Morgan (1977), Evans (1985), Campbell (1996, 2005), Noë (2004), Gallagher (2005), van Cleve (2007), Levin (2008a, 2008b), Glenney (2014, 2018), Vaughn (2019), Ferretti and Glenney (2021), Degenaar and Lokhorst (2021). For a recent review of experimental answers, see Spence and Stefano (2024). And there are much more.

² This interpretation and related issues can be traced back to Gallagher (2005), and have also been discussed in Levin (2008a), Glenney (2013), Ferretti (2018), and Cheng (2020). To continue using this terminology, the contrasting case is the "realist version," according to which Molyneux's question is treated as an empirical question that involves surgeries (Cheng, 2015; Connolly, 2013; Held et al., 2011; Schwenkler, 2012, 2013).

2 “Intrinsic similarity”

Readers might still wonder about the relation between our research questions and the Molyneux's question itself, so some more stage settings are in order. The Molyneux's question, no matter how one interprets it, is a yes-no question. It is easy enough to venture a positive or negative answer to the question, but what is more informative and important is the *rationale* behind the answer. For example, Berkeley (1975), Condillac (1930), and Lotze (1887) answered “no” due to their views about the heterogeneity of the senses. Thomson (1974), Evans (1985), and Noë (2004) answer “yes” due to common sensibles shared by sight and touch. Those rationales need to be flashed out in detail. Now, as anticipated above, I will argue that vision and touch generate distinct representation types, and the structural correspondence between these visual and tactile representations *explain* why I am inclined to answer “yes” to a specific version of Molyneux's question.

The “structural correspondence view” was introduced by E. J. Green (2022) as a critical target. He dubs his own picture the “type identity view.” In this section I will first characterise the dialectic as Green sees it, and then argue against Type identity. Section 3 will continue this negative case from a different angle, and argue for Structural correspondence.

To begin with, although Green does not use the same terminology, it seems that he is also focussing on the idealist version identified above, as he argues that real-world cases of “newly sighted subjects should not be relied on... because they are unlikely to form the kinds of shape representations responsible for cross-modal recognition in normal perceivers” (Green, 2022, pp. 694–695).³ Within this same framework, he focusses on this following issue of “whether there is an intrinsic similarity between our visual and tactual representations of shapes” (ibid., p. 695). He points out that if we opt for the positive answer, there are at least three ways to cash out “intrinsic similarity” in this context:

- a. **Type identity:** “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.” (ibid., p. 696)
- b. **Rational linkage:** “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

³ He does this *via* what he calls the “matching 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” (Green, 2022, p. 698). I shall not go into this in the paper, as in what follows I agree with Green that for his and our purposes, real cases from newly sighted subjects cannot settle the issues here.

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’).” (ibid., p. 696)⁴

- c. **Structural correspondence:** “vision and touch generate distinct representation type 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 (1981, bk. II, chap 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 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).” (Green, 2022, p. 696, original italics)⁵

Green’s descriptions of these views are quoted at length because they are quite accurate and informative. He also defines “intrinsic similarity” clearly: “the representations must closely match in their intrinsic characteristics – specifically their internal structure (e.g., imagistic or discursive format) and representational content” (ibid., p. 698).

Green then summarises his main argument in this succinct way: the empirical evidence he is going to cite “indicates that the visual and tactual representations are directly responsible for cross-modal recognition in normal perceivers exhibit both *physiological overlap* and *shared functional role*” (ibid., 707; original italics). He then invokes inference to the best explanation to support Type identity. We will briefly look into this case and see if Green’s argument is cogent.

From the context, it is clear that Green intends physiological overlap and shared functional role as complementary. He first points out that the lateral occipital complex (LOC) is the most likely candidate for sustaining viewpoint-invariant representations. This goes beyond a very weak claim that LOC is responsive to both seen and touched shapes. Erdogan, Chen, Garcea, Mahon, and Jacobs (2016)’s fMRI studies suggest that LOC is responsive to seen and touched shapes in similar ways.⁶ Now, Green concedes that this still falls short of establishing Type identity. Some pieces of behavioural evidence are in order.

⁴ Green mentions that Evans (1985) provides an influential defence of this view. For a charitable interpretation, see Salje (2019).

⁵ Green points out that there are more recent versions of this view, e.g., the idea that “shape properties are apprehended within a ‘tactile field’ that is distinct but analogous to the visual field (Cheng, 2019; Haggard & Giovagnoli, 2011)” (2022, p. 696).

⁶ Also see Tal and Amedi (2009) on repetition suppression in this context.

At this point, the physiological evidence is still neutral between the two following two hypotheses:

- a) Sight and touch produce tokens of *distinct* view-invariant representation types.
- b) Sight and touch produce tokens of *the same* view-invariant representation types.

Now Green argues that certain behavioural evidence concerning functional roles favours b). To begin with, “[t]here is a holistic, view-dependent stage, followed by a structural, view-invariant stage” (Green, 2022, p. 711). Bar (2001) shows that practice with multiple views might speed the perceptual formation of the relevant view-invariant representations, and such training-induced effects reflects important functional roles of view-invariant representations. Then, Lacey, Pappas, Kreps, Lee, and Sathian (2009) supports the idea that “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” (Green, 2022, p. 709). Here is where inference to the best explanation comes in: Green holds that the above considerations jointly support Type identity, the view that “[w]e visually recognize a previously touched object, and *vice versa*, because the very same representation types are produced in both cases” (ibid., p. 709).

The above is roughly Green’s positive case for Type identity. In the rest of this section and also next section, I will argue that his case for Type identity here is far from conclusive. What is crucial for Green is that “visual and tactual view-*invariant* representations are type-identical” (2022, p. 709, italics added). Now on the one hand, the role of visual and tactual view-*variant* or view-*dependent* representations seems to be underestimated (this section). On the other hand, even if we just focus on view-*invariant* representations, Structural correspondence is at least as plausible as Type identity (next section).

Historically, the relation between view-dependent and view-invariant properties are debated.⁷ This is especially salient in the context of viewing a tilted coin from various angles. Although there have been different formulations and construals in the past hundreds of years, the one fits Green’s overall framework is to ask if our perceptual system represents not only view-invariant properties (e.g., the coin’s objective shape) but also view-variant or view-dependent properties (e.g., the coin’s perspectival shapes). This age-old debate has regained its popularity in recent years. Morales, Bax, and Firestone (2020) conducted a series of experiments aiming to show that perspectival shapes are represented in the visual system. This has incurred several attacks, both conceptually and empirically (Burge & Burge, 2023; Linton, 2021). They have responded to some of the worries (Morales et al.,

⁷ Sometimes in the literature, the terminology is “viewpoint-dependent.” Here I follow Green’s usage. For a recent review of some aspects of this discussion, see Green and Schellenberg (2018).

2021; Morales & Firestone, 2023), and others have joined forces (Cheng et al., 2024). A convincing case has been made that perspectival shapes in particular, and view-dependent properties in general, are represented in the perceptual system. If this is the case, then it is unsatisfying to reach the conclusion of Type identity for view-invariant properties without taking into account of view-dependent properties. More specifically, these two kinds of properties behave differently both at the biological level and the functional level (Lin et al., 2024). Even if the pieces of empirical evidence cited by Green are all illuminating and useful, his interpretation seems to downplay the relevance of view-dependent properties. This is one reason why Green's abductive argument is less convincing than it might have appeared to be.

Now, one might respond on Green's behalf, that actually he grants that *both* sight and touch produce *both* view-dependent and view-invariant representations. According to this interpretation, his view is that the view-invariant representations are type identical, while the view-dependent representations are not. Given this, my above argument against Green might become much less powerful. The reason is that according to this interpretation, Green's argument *does* take into account view-dependent representations: the idea is that the *contrast* between patterns of orientation-dependence in intra-modal recognition and inter-modal recognition is best explained by appeal to a multi-tiered account, according to which the sensory modalities *initially* produce view-dependent representations, and *later* produce view-invariant representations. The view-dependent representations are usable *only* for intra-modal recognition, while the view-invariant representations are usable for *both* intra-modal and inter-modal recognition. If this is the apt interpretation of Green's picture, then it is fair to say that he does not downplay the importance of view-dependent representations; rather, it is more appropriate to understand him as holding that type identity arises at a different processing stage. Relatedly, this might imply that there is considerable scope for a compromise between Green's and my pictures, in this way: Structural correspondence might hold between view-dependent representations *within* vision and touch, while Type identity might hold *between* view-invariant representations.⁸ Therefore, it should be acknowledged that this sort of compromise position is available in the conceptual space.⁹

I leave the interpretative issues here for the readers to decide. Even if here I fail to show that Green underestimates the role of visual and tactual view-variant or view-dependent representations, as the above alternative interpretation suggests, there is still a case to be made for the idea that even if we just focus on view-invariant representations, Structural correspondence is at least as plausible as Type identity, *pace* Green.

⁸ See Green (2022), footnote 13, for textual evidence of this interpretation.

⁹ I thank a reviewer for this very detailed and charitable understanding of the dialectic.

3 Structural correspondence, memory, and imagination

What about we focus on view-invariant properties for now? The first half of this section will describe a version of Structural correspondence, and seek to show that it is at least as plausible as Type identity in the context of Molyneux's question.

As Green mentions, the tactile field hypothesis is a specific version of Structural correspondence, so it is totally possible to hold Structural correspondence without endorsing tactile fields. However, what will be developed here is *this* specific version of Structural correspondence based on tactile fields. It is natural to begin with sensory fields in general: visual fields are something people talk about in both daily and clinical contexts. Auditory fields (Wilson, 2023) and fields in other sensory modalities seem to be possible too. To be sure, we need relatively precise definitions to make progress, and in the case of touch, often one goes back to the one provided in Haggard and Giovagnoli (2011): tactile fields support "computation of spatial relations between individual stimulus locations, and thus [underlie] tactile pattern perception" (p. 65).¹⁰ One crucial qualification is that just like retinal images are *not* on any visual field, as the former is physiological while the latter is psychological, tactile fields are similar: they are *not* physiological receptive fields on human skins, even though psychological tactile fields and physiological receptive fields are certainly related. More positively, tactile fields are psychological *constructs* being postulated to explain and predict behavioural patterns, just like attention and working memory (Cheng, 2022).

The existence of tactile fields has been supported by a series of experiments (Fardo et al., 2018; Haggard & Giovagnoli, 2011; Serino et al., 2008) and theoretical discussions (Cheng, 2019, 2022; Cheng & Haggard, 2018). Although the exact nature of them is still debated (Skrzypulec, 2021, 2022), it would require substantive moves to deny their existence. From Green's perspective, Type identity offers a better explanation. But there are at least two problems here. First, it is always tricky to compare hypotheses concerning simplicity and explanatory power, etc. It is true that "tactile fields" are additional postulates, but their existence explains and predicts tactile pattern perceptions (see the papers cited in this paragraph). Secondly, Type identity might make sight and touch too similar. Green acknowledges that others have pointed to *dissimilar* aspects between the two senses (he mentions (Martin, 1992; O'Shaughnessy, 1989; Prinz, 2002; but also see Soteriou, 2013); however, he does not go into details of those aspects. If sight and touch deliver exactly the same type of representations, it is unclear how such substantive dissimilarities can be accommodated. Structural correspondence, by contrast, captures intrinsic similarities between the two senses, while respects their disanalogies.

To see this, we do need to look into some details of the dissimilar aspects. Although O'Shaughnessy, Martin, Prinz, and Soteriou all hold slightly different

¹⁰ Papers from the Haggard team tends to use the singular "a tactile field," while I always prefer plurals here, as there are multiple tactile fields exemplified by a single organism.

views, I will use Martin's discussion as our prime example. In what follows I will briefly summarise some key points in his 1992 paper, and explain why we should favour Structural correspondence over Type identity if we take those points seriously.

Martin begins with the observation that though obviously we can identify shapes and sizes either by sight or by touch, both their physiological and conscious characters are very different. This situation generates at least two questions. The first is in the affinity of our Molyneux's question: how it is that the same properties can be perceived by the two senses? Martin rather focusses on this second question: "given that the same properties are perceived, where does the difference between the senses lie" (Martin, 1992, p. 196). Following O'Shaughnessy (1989), Martin goes for a Berkeleyan solution, i.e., to look for *structural differences*. From here, one might naturally think that this approach is outright incompatible with Structural correspondence. Indeed, this is how the dominant interpretation goes. After characterising Martin's main ideas, I will explain how Structural correspondence sits well with this Berkeleyan solution.

The most quoted passage in this neighborhood is probably this:

[T]he visual field plays a role in sight which is not played by any sense field in touch. Touch is dependent on bodily awareness and if, or where, that involve a sense field, it does so in a strikingly different way from that in which visual experience involves the visual field. (Martin, 1992, p. 197)

One crucial point here is how we should interpret "strikingly different." We will explain this presently. Note that "visual fields" in Berkeley and O'Shaughnessy are tied to sense-datum theories, but talks about sensory fields can be detached from such specific theories of perception. Now, normal visual experiences are not only of objects that are located in some space, but also as of an overarching space in which those objects reside. By contrast, tactile experiences seem to be strikingly different, to use Martin's phrase. When one tells the shape of a glass, for example, vision figures out the answer rather quickly and directly, while touch takes much longer time with haptic explorations that crucially depend on bodily awareness, i.e., ways in which we are aware of our own bodies, including proprioception, kinaesthesia, and equilibrioception, etc. On this view, the body is like a *template*: one uses one's own body to measure the shapes and sizes of the external objects in question.

The above summary does not do justice to Martin's rich paper, but it is minimally enough for our purposes. The crucial question here is: whether Structural correspondence fares better than Type identity with regard to this Berkeleyan picture. Why does Martin's observations do not clash with Structural correspondence, contra the standard interpretation? Recall his remark about how sight and touch are strikingly different: we can respect the thesis that touch essentially depends on bodily awareness, and this is how spatial sight and touch differ. However, this does not mean that tactile experiences are devoid of spatial contents independent

of bodily awareness. Rather, as the papers from the Haggard team shows, the skin space (physiological basis) that hosts tactile fields (psychological construct) exemplifies basic spatial contents that are not provided by bodily awareness. The upshot is this: although it is true that there are structural *differences* between sight and touch, it does not preclude the idea that there are also structural *correspondences* between the two senses.

But compatibility is not enough. Is it really the case that Structural correspondence fares better than Type identity in light of the above Berkeleyan picture? Recall what Type identity says: vision and touch generate tokens of the same representation type. It is notoriously difficult to spell out the type/token distinction, but normal examples will do. For example, "MacBook air 13-inch 2024" is a type, which can have various tokens that appear in different locations at different times. Those tokens are not exactly the same: they are not numerically identical. However, many of them are so similar that each one can replace one another without being noticed. Can this be done between sight and touch? It seems unlikely. On the one hand, given Martin's observations, how can sight and touch deliver the same type of spatial representations? On the other hand, even without those observations, it is unclear the similarity between the two senses is so high that type identity is guaranteed. Type identity just seems too strong.

One might respond on Green's behalf (again) that on his view, Type identity only holds for one amongst many spatial representations produced by vision and touch. Therefore, one cannot simply replace or swap the *entirety* of visual and tactile spatial representations with each other, since there are also many inter-modal differences in spatial representations, for example at the level of view-dependent representations. At best, one could replace or swap *one aspect* of spatial representations in vision and touch, namely their object-centred shape contents. If so, my argument above seems to be too quick.¹¹ This is indeed a conceptual possibility, but it seems gerrymandering, and short of empirical support: how do we *individuate* aspects of spatial representations? And even if we have a non-arbitrary, principled way to single out object-centred shape contents, how do we know *empirically* that it is this aspect that is being replaced or swapped? To be sure, these challenges might be answerable by Green and others, but here I put these challenges on the table for further investigations.

At this point, it is useful to consider the relation between view-dependent/view-invariant representations on the one hand, and tactile field representations on the other. Are tactile field representations view-dependent or not? Here, I tentatively regard those representations as *view-invariant*. But is there any rationale behind this move? Indeed, some might think that tactile field representations are *view-dependent*. The reason is that an object's position in the tactile field changes with changes in its position and orientation relative to the perceiver's body.¹² Now here is why tactile field representations are likely to be view-invariant: as indicated

¹¹ I thank a reviewer for pressing me on this.

¹² Again, I thank a reviewer for this incisive potential objection.

above, tactile fields are postulated to explain and predict tactile pattern perceptions. More specifically, it is supposed to explain why the subject can reach a *stable* representation of a given shape (say, “this is roughly square”; its *objective shape*) *via* the varieties of haptic angles (its *perspectival shape*). It is true that in the course of explorations, view-dependent representations are at work. But the key point is that tactile field representations are the *end products* of such explorations, and those products are themselves stable, view-invariant representations of *objective shapes*. Why should we expect tactile field representations (given their structure) to remain the same across changes in an object’s orientation? Compare visual field representations: it is true that there will be view-dependent representations as we move or the objects move, but it is also true that whatever the movements are, there will be stable end products, such as “that coin just looks round no matter how we view it.”

One might wish to know more about what the structural correspondence between the visual field and the tactile field representations consists in.¹³ Despite being abstract and speculative, I propose that there might be *spatial isomorphisms* between the relevant visual and tactile spatial representations. Suppose that a subject both sees and touches a specific object, and represents it as occupying certain portions of both the visual field and the tactile field. The potential spatial isomorphisms include corresponding geometries, even gestalts (Cataldo et al., *in preparation*). The structural correspondence crucially contributes to the presumed “yes” answer to this version of Molyneux’s question.

Now, assuming that we have made plausible the idea that Structural correspondence – more specifically, the version involving the analogy between visual fields and tactile fields – provides part of the *right* answer for the idealist version of Molyneux’s question, a follow-up question naturally suggests itself: is it the *full* answer? Since the question is about sight and touch, the full answer should include a good characterisation of the relevant aspects of sight and touch for sure. But that does not seem to be enough. After all, the human subjects need to make judgements and answer the question based on those judgements. In order to make the relevant judgements, however, inputs from perceptions are not enough. The rest of this section will say a bit about in what ways memory and imagination have to be involved in such process.

Again, in the idealist version of Molyneux’s question, the subjects are made to see right away by perfect surgeries or procedures, which guarantee that their visual capacities are as good as they can get as long as they are within the standard range of human vision. The structural correspondence view has it that spatial sight and spatial touch, though different in many ways for sure, nevertheless share crucial spatial structures identified by the tactile field hypothesis. Now, even with these two resources at hand – good enough vision and close enough spatial structures – the subjects still cannot make the relevant correct judgements, because

¹³ This is yet another query from a reviewer.

they need to *remember* the feelings of touching those shapes, and *imagine* what it would be like to see those shapes. These need to be cashed out.

Let's start with memory. Before Molyneux's subjects have their crucial surgeries, they have touched the relevant objects as much as they wish. They *tactilely remember* those shapes. During the task, they are not allowed to touch those target objects. That means the kind of memory involved is essential; the retention of tactile memories is the precondition for proceeding in the task. What kind of memory? In vision experiments, stimuli often appear very briefly and then disappear. In such cases, visual working memory is involved. In the Molyneux's case, it is *tactile working memory* that is relevant, as it is a visual task that relies on memories of tactile experiences. The topography (Harris et al., 2001) and neural correlates (Esmaeili & Diamond, 2019) of tactile working memory have been well studied. Moreover, these memories must be *multisensory* in that they need to be shared between sight and touch (Quak et al., 2015): the relevant touch-based visual information is required to make the subsequent judgements.

What about imagination? It might be less obvious that imagination is involved in the Molyneux's task. After all, at no stage the subjects need to *imagine* the shapes, it might seem. Contrary to the appearance, however, imagination plays some important roles for various reasons. For example, historically speaking, there has been a long tradition for the idea that imagination accompanies all perceptions (Brown, 2018; Kant, 1929; Strawson, 1982). But surely this is not enough: on the one hand, this thesis is not entirely uncontroversial and might depend on what one means by "imagination" here¹⁴; on the other hand, it is a general thesis on perception, so not specific enough for the Molyneux's case. However, during the task, the subjects need to imagine *what it would be like if I touch the target objects*. In doing so, *multisensory imagination* is required (Krüger et al., 2022), as their imagination and subsequent judgements are *based* on tactile experiences but *about* expectations of visual experiences. After the task, the subjects might think that "well, that was different from what I have *imagined* in such and such respects." We can draw two tentative conclusions here: first, even if Structural correspondence explains intrinsic similarities between sight and touch, it is far from enough to explain and predict the behaviours of Molyneux's subjects; both memory and imagination are required. Secondly, and relatedly, Molyneux's subjects are not doing a *perceptual* task only; it is also a task about *memory* and *imagination*. This might mean that to satisfactorily address Molyneux's question, at least certain versions of it, integral considerations concerning perception, memory, and imagination are required.

There is another important aspect of imagination that connects us naturally to the concluding section. Imagination, when properly used, can be a source of *knowledge of possibilities*, or *modal knowledge* (Liao & Gendler, 2019). As indicated above, for Molyneux's subjects, they will need to imagine what it would be like if I touch

¹⁴ E.g., the capacity for forming novel representations vs. certain fictional attitudes, in (van Leeuwen, 2013); spontaneous vs. deliberate, occurrent vs. non-occurrent, social vs. solitary, in (Walton, 1990); also see (Strawson, 1982), on image, imagine, and imagination.

the target objects. In doing so, they might thereby gain modal knowledge about their potential, future tactile experiences. To be sure, such prediction can be confirmed or disconfirmed. In the latter case, the predictions or anticipations would not be knowledge. But sometimes, those predictions *can be* knowledge. There have been many discussions of this and related issues under the label of “conceivability and possibility” (Gendler & Hawthorne, 2002). More commonly, it is about thought experiments such as those concerning philosophical zombies and physicalism (Chalmers, 2002). Given that the present version of Molyneux’s question is a thought experiment too (the “idealist version”), the connection to imagination brings us to the realm of modal knowledge. Although conceivability is not exactly the same as imaginability (Fiocco, 2007), they do have much overlap. Is it conceivable that such and such visual shapes *would feel like this*, when touched? It is the kind of imagination and modal knowledge the Molyneux’s subjects are supposed to be engaging.

4 Conclusion: Multisensory Knowing-How

Now trivially, what Molyneux’s subjects are facing is a multisensory or crossmodal task.¹⁵ However, to get clear about the multisensory nature of it is no trivial task. Although not usually framed as an epistemological question, from the end of the previous section it should be clear that Molyneux’s question is actually about *knowing* as well: do the subjects *know* which one is the globe, and which one is a cube? Crucially, it is *knowing-how* because the Molyneux’s task is a practical challenge for the subjects: they are not supposed to solve or think of the problem on site in any theoretical way. This point holds quite independently of which view of knowing-how one prefers. Even if one endorses the intellectualist picture here (*pace* Ryle, 1946, 1949; and Noë, 2005) – roughly the idea that knowing-how is in one way or another reducible to propositional knowledge (e.g., Snowdon, 2004; Stanley, 2011; Stanley & Williamson, 2001) – it is still true that certain kind of knowledge-how is in play. Note that here we do not preclude the possibility that the subjects also know *that* the visible object is a sphere or a cube. However, arguably this knowing-that is derivative and abstracted from practical knowledge. The Molyneux’s subjects need to step back and reflect to gain such a propositional knowledge. In both knowing-how and knowing-that here, they are seeking to gain *knowledge through imagination* (Kind & Kung, 2016), *with the help of memory and Structural correspondence*. Now, to give more positive and empirical contents of this “multisensory knowing-how” view, one needs to figure out *spatially*, how the same shape could be felt alike in sight and in touch, and this is by and large taken care of by Structural correspondence. However, the *temporal* aspect seems to be more challenging: given that sight tends to be much faster when functioning normally,

¹⁵ These two terms do not mean exactly the same thing in the literature, but this does not affect the points and argumentation of this paper.

how could Molyneux's subjects compensate this temporal difference practically in their imaginations (Salje, 2019)? This will be a crucial question for further research.

References

- Bar, M. (2001). Viewpoint dependency in visual object recognition does not necessarily imply viewer-centered representation. *Journal of Cognitive Neuroscience*, 13, 793–799. <https://doi.org/10.1162/08989290152541458>
- Berkeley, G. (1975). *An essay toward a new theory of vision (1709)* (M. R. Ayers, Ed.). J. M. Dent.
- Brown, D. H. (2018). Infusing perception with imagination. In F. Macpherson & F. Dorsch (Eds.), *Perceptual imagination and perceptual memory* (pp. 133–160). Oxford University Press.
- Burge, J., & Burge, T. (2023). Shape, perspective, and what is and is not perceived: Comment on Morales, Bax, and Firestone (2020). *Psychological Review*, 130(4), 1125–1136. <https://doi.org/10.1037/rev0000363>
- Campbell, J. (1996). Molyneux's question. *Philosophical Issues*, 7, 301–318. <https://doi.org/10.2307/1522914>
- Campbell, J. (2005). Information-processing, phenomenal consciousness and Molyneux's question. In J. L. Bermúdez (Ed.), *Thought, reference, and experience: Themes from the philosophy of Gareth Evans*. Clarendon Press.
- Cataldo, A., Cheng, T., Schwenkler, J., & Haggard, P. (in preparation). *Constructing the tactile field: Somatosensory gearing for spatial structures*.
- Chalmers, D. J. (2002). Does conceivability entail possibility? In T. S. Gendler & J. Hawthorne (Eds.), *Conceivability and possibility* (pp. 145–200). Oxford University Press.
- Cheng, T. (2015). Obstacles to testing Molyneux's question empirically. *I-Perception*, 6(4), 1–5. <https://doi.org/10.1177/2041669515599330>
- 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* (pp. 226–247). Routledge.
- Cheng, T. (2020). Molyneux's question and somatosensory spaces. In G. Ferretti & B. Glenney (Eds.), *Molyneux's question and the history of philosophy*. Routledge.
- Cheng, T. (2022). Spatial representation in sensory modalities. *Mind and Language*, 37(3), 485–500. <https://doi.org/10.1111/mila.12409>
- Cheng, T., & Haggard, P. (2018). The recurrent model of bodily spatial phenomenology. *Journal of Consciousness Studies*, 25(3-4), 55–70.
- Cheng, T., Lin, Y., & Wu, C.-W. (2024). Perspectival shapes are viewpoint-dependent relational properties. *Psychological Review*, 131(1), 307–310. <https://doi.org/10.1037/rev0000404>
- Condillac, E. B. (1930). *Condillac's treatise on the sensations (1754)* (G. Carr, Trans.). Favil Press.
- Connolly, K. (2013). How to test Molyneux's question empirically. *I-Perception*, 4(8), 508–510. <https://doi.org/10.1068/i0623jc>
- Connolly, K. (2019). *Perceptual learning: The flexibility of the senses*. Oxford University Press. <https://doi.org/10.1093/oso/9780190662899.001.0001>
- Degenaar, M., & Lokhorst, G. J. (2021). *Molyneux's problem* (E. N. Zalta & U. Nodelman, Eds.). The Stanford Encyclopedia of Philosophy. <https://plato.stanford.edu/archives/sum2024/entries/molyneux-problem/>
- 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. https://doi.org/10.1162/jocn_a_00937
- Esmaeili, V., & Diamond, M. E. (2019). Neuronal correlates of tactile working memory in prefrontal and vibrissal somatosensory cortex. *Cell Reports*, 27, 3167–3181. <https://doi.org/10.1016/j.celrep.2019.05.034>
- Evans, G. (1985). *Molyneux's question*. In *his collected papers*. Oxford University Press.
- Fardo, F., Beck, B., Cheng, T., & Haggard, P. (2018). A mechanism for spatial perception on human skin. *Cognition*, 178, 236–243. <https://doi.org/10.1016/j.cognition.2018.05.024>
- Ferretti, G. (2018). Two visual systems in Molyneux's subjects. *Phenomenology and the Cognitive Sciences*, 17, 643–679. <https://doi.org/10.1007/s11097-017-9533-z>
- Ferretti, G., & Glenney, B. (Eds.). (2021). *Molyneux's question and the history of philosophy*. Routledge.
- Fiocco, M. O. (2007). Conceivability, imagination, and modal knowledge. *Philosophy and Phenomenological Research*, 74(2), 367–380. <https://doi.org/10.1111/j.1933-1592.2007.00022.x>
- Gallagher, S. (2005). *How the body shapes the mind*. Oxford University Press.
- Gendler, T. S., & Hawthorne, J. (2002). *Conceivability and possibility*. Oxford University Press.
- Glenney, B. (2013). Philosophical problems, cluster concepts, and the many lives of Molyneux's question. *Biology and Philosophy*, 28(3), 541–558. <https://doi.org/10.1007/s10539-012-9355-x>
- Glenney, B. (2014). *Molyneux's question*. <https://iep.utm.edu/molyneux/#:~:text=Molyneux's%20question%2C%20also%20known%20as,to%20John%20Locke%20in%201688>
- Glenney, B. (2018). Molyneux's problem. In E. Craig (Ed.), *Routledge encyclopedia of philosophy*. Routledge.
- Green, E. J. (2022). Representing shape in sight and touch. *Mind and Language*, 37(4), 694–714. <https://doi.org/10.1111/mila.12352>
- Green, E. J., & Schellenberg, S. (2018). Spatial perception: The perspectival aspect of perception. *Philosophy Compass*, 13(2), e12472. <https://doi.org/10.1111/phc3.12472>
- Haggard, P., & Giovagnoli, G. (2011). Spatial patterns in tactile perception: Is there a tactile field? *Acta Psychologica*, 137(1), 65–75. <https://doi.org/10.1016/j.actpsy.2011.03.001>
- Harris, J. A., Harris, I. M., & Diamond, M. E. (2001). The topography of tactile working memory. *Journal of Neuroscience*, 21(20), 8262–8269. <https://doi.org/10.1523/JNEUROSCI.21-20-08262.2001>

Cheng, T. (2024). Structural correspondence in Molyneux's subjects. *Philosophy and the Mind Sciences*, 5. <https://doi.org/10.33735/phimisci.2024.11572>



- Held, R., Ostrovsky, Y., Gelder, B. de, Gandhi, T., Ganesh, S., Mathur, U., & Sinha, P. (2011). The newly sighted fail to match seen with felt. *Nature Neuroscience*, *14*(5), 551–553. <https://doi.org/10.1038/nn.2795>
- Kant, I. (1929). *Critique of pure reason (1781-1787)* (N. Kemp Smith, Trans.). Macmillan.
- Kind, A., & Kung, P. (2016). *Knowledge through imagination*. Oxford University Press.
- Krüger, B., Hegele, M., & Rieger, M. (2022). The multisensory nature of human action imagery. *Psychological Research*, 1870–1882. <https://doi.org/10.1007/s00426-022-01771-y>
- 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. <https://doi.org/10.1007/s00221-009-1856-8>
- Leibniz, G. W. (1981). *New essays on human understanding (1765)* (P. R. Remnant & J. Bennett, Eds.). Cambridge University Press.
- Levin, J. (2008a). Molyneux meets Euthyphro. *Croatian Journal of Philosophy*, *8*(3), 289–297.
- Levin, J. (2008b). Molyneux’s question and the individuation of perceptual concepts. *Philosophical Studies*, *139*(1), 1–28. <https://doi.org/10.1007/s11098-007-9072-5>
- Liao, S.-Y., & Gendler, T. (2019). Imagination. *The Stanford Encyclopedia of Philosophy*. <https://plato.stanford.edu/archives/sum2020/entries/imagination/>
- Lin, Y., Hsu, Y.-Y., Cheng, T., Hsiung, P.-C., Wu, C.-W., & Hsieh, P.-J. (2024). *Neural representations of perspectival shapes and attentional effects: Evidence from fMRI and MEG*. *Cortex*. <https://doi.org/10.1016/j.cortex.2024.04.003>
- Linton, P. (2021). Conflicting shape percepts explained by perception cognition distinction. *Proceedings of the National Academy of Sciences*, *118*(10), e2024195118. <https://doi.org/10.1073/pnas.2024195118>
- Locke, J. (1975). *An essay concerning human understanding (1694)* (P. H. Nidditch, Ed.). Oxford University Press.
- Lotze, H. (1887). *Metaphysic* (B. Mosanquet, Ed.). Clarendon Press.
- Macpherson, F., & Dorsch, F. (2018). *Perceptual imagination and perceptual memory*. Oxford University Press.
- Martin, M. G. F. (1992). Sight and touch. In T. Crane (Ed.), *The contents of experience*. Cambridge University Press.
- Morales, J., Bax, A., & Firestone, C. (2020). Sustained representation of perspectival shape. *Proceedings of the National Academy of Sciences*, *117*(26), 14873–14882. <https://doi.org/10.1073/pnas.2000715117>
- Morales, J., Bax, A., & Firestone, C. (2021). Reply to Linton: Perspectival interference up close. *Proceedings of the National Academy of Sciences*, *118*(28), e2025440118. <https://doi.org/10.1073/pnas.2025440118>
- Morales, J., & Firestone, C. (2023). *Empirical evidence for perspectival similarity*. *Psychological Review*. <https://doi.org/10.1037/rev0000403>
- Morgan, M. J. (1977). *Molyneux’s question: Vision, touch, and the philosophy of perception*. Cambridge University Press.
- Noë, A. (2004). *Action in perception*. MIT Press.
- Noë, A. (2005). Against intellectualism. *Analysis*, *65*(4), 278–290. <https://doi.org/10.1093/analys/65.4.278>
- O’Shaughnessy, B. (1989). The sense of touch. *Australasian Journal of Philosophy*, *67*(1), 37–58. <https://doi.org/10.1080/00048408912343671>
- Prinz, J. (2002). *Furnishing the mind: Concepts and their perceptual basis*. MIT Press.
- Quak, M., London, R. E., & Talsma, D. (2015). A multisensory perspective of working memory. *Frontiers in Human Neuroscience*, *9*, 197.
- Ryle, G. (1946). Knowing how and knowing that: The presidential address. *Proceedings of the Aristotelian Society*, *46*(1), 1–16.
- Ryle, G. (1949). *The concept of mind*. University of Chicago Press.
- Salje, L. (2019). The inside-out binding problem. In T. Cheng, O. Deroy, & C. Spence (Eds.), *Spatial senses: Philosophy of perception in an age of science*. Routledge.
- Schwenkler, J. (2012). On the matching of seen and felt shapes by newly sighted subjects. *I-Perception*, *3*, 186–188. <https://doi.org/10.1068/i0525ic>
- Schwenkler, J. (2013). Do things look the way they feel? *Analysis*, *73*, 86–96.
- Serino, A., Giovagnoli, G., Vignemont, F. de, & Haggard, P. (2008). Spatial organisation in passive tactile perception: Is there a tactile field? *Acta Psychologica*, *128*(2), 355–360. <https://doi.org/10.1016/j.actpsy.2008.03.013>
- Skrzypulec, B. (2021). Spatial content of painful sensations. *Mind and Language*, *36*(4), 554–569. <https://doi.org/10.1111/mila.12358>
- Skrzypulec, B. (2022). Is there a tactile field? *Philosophical Psychology*, *35*(3), 301–326. <https://doi.org/10.1080/09515089.2021.1980519>
- Snowdon, P. F. (2004). Knowing how and knowing that: A distinction reconsidered. *Proceedings of the Aristotelian Society*, *104*(1), 1–29. <https://doi.org/10.1111/J.0066-7373.2004.00079.X>
- Soteriou, M. (2013). *The mind’s construction: The ontology of mind and mental action*. Oxford University Press.
- Spence, C., & Stefano, N. D. (2024). Old and new versions of the Molyneux question: A review of experimental results. *Philosophy and the Mind Sciences*, *5*. <https://doi.org/10.33735/phimisci.2024.11337>
- Stanley, J. (2011). *Know how*. Oxford University Press.
- Stanley, J., & Williamson, T. (2001). Knowing how. *Journal of Philosophy*, *98*(8), 411–444.
- Strawson, P. F. (1982). Imagination and perception (1970). In R. C. S. Walker (Ed.), *Kant on pure reason*. 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. <https://doi.org/10.1007/s00221-009-1949-4>
- Thomson, J. (1974). Molyneux’s problem. *Journal of Philosophy*, *71*(18), 637–650. <https://doi.org/10.2307/2024801>
- van Cleve, J. (2007). Reid’s answer to Molyneux’s question. *The Monist*, *90*(2), 251–270.
- van Leeuwen, N. (2013). The meaning of “imagine” part I: Constructive imagination. *Philosophical Compass*, *8*(3), 220–230. <https://doi.org/10.1111/j.1747-9991.2012.00508.x>

Cheng, T. (2024). Structural correspondence in Molyneux’s subjects. *Philosophy and the Mind Sciences*, *5*. <https://doi.org/10.33735/phimisci.2024.11572>



©The author(s). <https://philosophymindscience.org> ISSN: 2699-0369

- Vaughn, A. (2019). Is Locke's answer to Molyneux's question inconsistent? Cross-modal recognition and the sight-recognition error. *Canadian Journal of Philosophy*, 49(5), 670–688. <https://doi.org/10.1080/00455091.2018.1444899>
- Walton, K. L. (1990). *Mimesis as make-believe*. Harvard University Press.
- Wilson, K. A. (2023). The auditory field: The spatial character of auditory experience. *Ergo: An Open Access Journal of Philosophy*, 9(40), 1080–1106. <https://doi.org/10.3998/ergo.2909>

Open Access

This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

