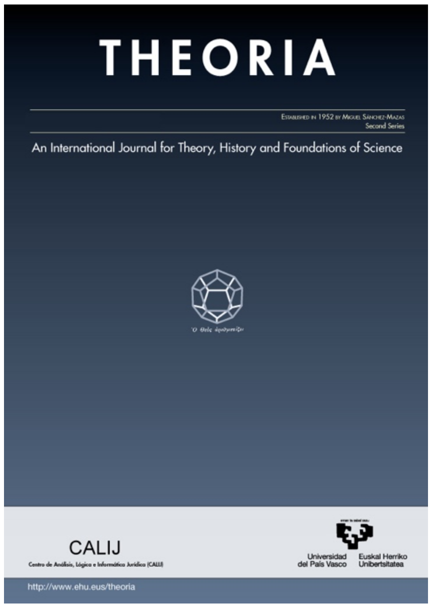
JOURNAL PRE-PROOF



Perceptual reference, object files and Molyneux's Question

Sergio Cermeño-Aínsa

DOI: 10.1387/theoria.27048

Received: 03/11/2024

Final version: 16/07/2025

PERCEPTUAL REFERENCE, OBJECT FILES AND MOLYNEUX'S QUESTION

(Referencia perceptual, archivos de objetos y la cuestión de Molyneux)

Sergio Cermeño-Aínsa
University of Girona

ABSTRACT: Molyneux's Question (MQ) —whether a newly sighted might immediately recognize tactilely familiar shapes by sight alone— remains inconclusive. The most common way to pose the question is in representational terms, i.e., whether vision and touch generate or not similar representational types or whether there is some intrinsic similarity between visual and tactual shape representations. Recent developments in cross-modal perception suggest that if visual and tactile representations are linked, then a positive response to MQ is the most probable conclusion. In this paper, I explore and provide some suggestions for this possibility. In particular, I give rise to the object file strategy, according to which, if part of the tactile information is referentially encoded and amodally stored in object files, then the category-specific contents of R(vision) and R(touch) are in some way propositionally (conceptually) linked, thus opening a wide avenue to think that a newly sighted might recognize tactilely familiar shapes by sight alone.

Keywords: Molyneux Question; Perceptual Shape Representation; Cross-modal Perception; Object Files.

RESUMEN: La Cuestión de Molyneux (MQ por sus siglas en inglés) —si un recién vidente podría reconocer visualmente formas previamente conocidas solo mediante el tacto— sigue sin ser concluyente. La forma más común de plantear esta cuestión es en términos representacionales, es decir, si la visión y el tacto generan tipos de representación similares o si existe alguna similitud intrínseca entre las representaciones visuales y táctiles en el reconocimiento de las formas. Avances recientes en la percepción intermodal sugieren que, si las representaciones visuales y táctiles están vinculadas, una respuesta positiva a MQ es la más probable. En este artículo, exploro y presento algunas ideas que sugieren esta posibilidad. En particular, planteo la estrategia de los archivos de objetos, según la cual, si parte de la información táctil se codifica referencialmente y se almacena amodalmente en archivos de objetos, entonces los contenidos específicos de la categoría de R(visión) y R(tacto) estarán de alguna manera vinculados proposicionalmente (conceptualmente). Esto abre una amplia posibilidad para pensar que una persona recién vidente podría reconocer visualmente formas previamente conocidas solo mediante el tacto.

Palabras Clave: Cuestión de Molyneux; Representación Perceptual de las Formas; Percepción Intermodal; Archivos de Objetos.

SHORT SUMMARY: This paper defends a solution to the Molyneux Question grounded on what I call the *object files strategy*. Roughly, if tactile information is referentially encoded and amodally stored in object files, then the category-specific contents of R(vision) and R(touch) are in some way propositionally (conceptually) linked, and therefore a newly sighted might recognize tactilely familiar shapes by sight alone.

1. *Introduction*

Molyneux's Question (MQ)[1] asks whether a person who has recently gained the ability to see would be able to recognize familiar tactile shapes, such as spheres or cubes, by sight alone. There are three possible responses: negative, positive and plural. Negative answers usually appeal to perceptual learning; simply, learning to see a cube is fundamentally different from learning to feel a cube. This is because the connections between our ideas of different shapes (cube and sphere) acquired through different sense modalities must be learned through direct experience; i.e., sensory ideas of sight and touch are inherently different (Locke, 1694; Berkeley, 1709; Condillac, 1754; Lotze, 1887). Arguably, negative responses have dominated the scene, and it is perhaps the first impression for everyone who has been in touch with MQ. Furthermore, negative responses have been supported, though not firmly, by empirical research.[2] As such, the burden of proof falls on those who argue for an affirmative answer.

Therefore, the challenge lies more in how to stand up for a compelling positive answer than in offering more arguments for the negative one. Affirmative answers have usually appealed to some kind of nativism concerning the mechanisms for conceptual acquisition. Different arguments have tried to substantiate the affirmative option. The first and most general suggests that subjects imaginatively construct a visual representation from their previous tactile experience (Leibniz, 1760). Current arguments, however, are more specific and amply supported by progress in developmental science and neuroscience. Noë (2004), for example, thinks that sensorimotor contingencies in sight and touch can share a common spatial content. Campbell (1996) appeals to the geometrical properties of objects, which constitute subjects' sensory experience, and makes the perception of shape by sight and touch similar in structure. Evans (1985) argues that our perception of the sphere and the cube are the same for touch and sight because both share the same egocentric relations, i.e., the ability to egocentrically localize the parts of shapes is alien to the sense modality in which it is perceived. For Held (2009), there must be a transference from tactile to visual representations via supra-modal pathways. And similarly, Pascual-Leone and Hamilton (2001) talk about meta-modal brain centres that perform particular computational operations without specific reference to any type of sensory modality.[3]

Beyond the yes/no responses there is still a third possibility. It has also been proposed that a monolithic answer to MQ is too simplistic; the complexity of MQ involves more than one problem and, therefore, requires more than one solution. This is the pluralist view. Reid (1764), for example, argued that MQ is ambiguous since a positive or negative response depends on whether the shapes are two or three-dimensional as well as on the expertise of the subjects. Matthen and Cohen (2020) claim that MQ comprises a cluster of subproblems, each of which requires a particular solution. And Nanay (2020) thinks that a compelling answer is only possible by considering the specificities

---

1 For a recent recompilation of essays exploring the long-standing issues that MQ presents to philosophers and psychologists, see Ferreti and Gleney (2021).

2 The first experimental data concerning MQ dates back to Cheselden (1728), who published an account of what a congenitally blind person had seen after his cataracts were removed. More recently, there have been several attempts to approach the issue empirically, such as Gregory and Wallace (1963), Fine et al. (2003), and especially Held et al. (2011), who conducted what has been the most discussed experiment to date. As we will see, this evidence is not free of controversies.

3 For other positive arguments, see Prinz (2002) or Levin (2008).

of the type of blindness in question. In particular, whether blindness involves or not the impairment of visual mental imagery.

In this paper, I argue that it is possible to defend an affirmative response by employing what I call the object file strategy (OFS). OFS is consistent with some of the positive answers reviewed above; it shares many things with what Held (2009) called transference from tactile to visual representations via supra-modal pathways, and it is markedly aligned with what Pascual-Leone and Hamilton (2001) refer to as meta-modal brain centres that perform particular computational operations without specific reference to any type of sensory modality. However, as we will see, OFS has its proper specificities.

But let's start from the beginning. Any affirmative response to MQ requires understanding seeing and touching in representational terms. The question is: Do vision and touch generate (or not) similar representational types? Put differently, is there an intrinsic similarity, a common code, between visual and tactual shape representations? Suppose F is a shape property represented by R. Three possibilities deploy here:

1. If R(vision) and R(touch) are *identical*, then a newly sighted might recognize tactilely familiar shapes by sight alone.
2. If R(vision) and R(touch) are *referentially linked*, then a newly sighted might recognize tactilely familiar shapes by sight alone.
3. If R(vision) and R(touch) are *structurally linked*, then a newly sighted might recognize tactilely familiar shapes by sight alone.[4]

These are some of the possible positive responses defended by theorists. Not many see it as viable defending (1); there is no easy way to support that R(vision) and R(touch) produce the same types of representations. I think this option depends on what we understand by identical representations; basically on whether these similarities are in perceptual shape representations or recognitional shape representations. The latter may be identical since the concept of 'sphere' is the same for any sense modality, but this is not necessarily the case for the former (for discussion, see Green, 2022). Many others defend different versions of (2) or (3). One version of (2) was initially advocated by Evans (1985). Particularly, Evans assumes that if "the tactual concept is the same as the visual concept", then the answer to MQ is 'yes' (p. 381). And versions of (3) are recently supported by Cheng (2019). Option (3) depends on the existence of a tactile field structurally similar to the visual field (Cheng, 2019; Serrahima, 2023). However, there is no easy way to explain that both fields (tactual and visual) are identical beyond their referential link. For example, just as tactile experiences of external objects are mediated by interoceptive bodily sensations, visual experiences are purely exteroceptive (Martin, 1993; O'Shaughnessy, 1989). So, I propose to go on with (2) here. Option (2) puts the accent on a referential link between the two representations, but it does not require that the representations are identical or structurally linked. Specifically, I suggest that by leaning on a particular characterization of the notion of object file, it is possible to take a step toward a plausible variant of (2). My core claim reads as follows:

[4] Of course, if R(vision) and R(touch) are not identical and do not have any conceptual or structural link, then a newly sighted will not recognize tactilely familiar shapes by sight alone.

Object Files Strategy: If part of the tactile information is *referentially* encoded and *amodally* stored in object files, then the *category-specific* contents of R(vision) and R(touch) are in some way *propositionally (conceptually) linked*, and therefore a newly sighted might recognize tactilely familiar shapes by sight alone.

OFS admits a difference between tactile and visual representation —they are not necessarily similar nor structurally linked but only referentially bonded. This makes OFS more in line with option (2).[5] In what follows, I will defend this position.

The paper is structured as follows. Section 2 provides some clarifications, specifications and particularities of MQ. This preliminary section pretends to clarify the nature of MQ and its peculiarities. Section 3 critically reviews some classical and novel experimental attempts to solve MQ. Section 4 focuses on recent developments in cross-modal perception. I argue that any positive answer to MQ requires (though does not exhaust) the existence of cross-modal perception. Section 5 presents and defends Object Files as the mechanisms for harbour cross-modal perceptual recognition. Section 6 deploys the Object File Strategy by presenting the properties of object files: referentially (subsection 6.1), amodality (subsection 6.2), propositionally (subsection 6.3) and category-specifically (subsection 6.4), and defends that, thus seen, OFS is a plausible view. Section 7 offers empirical evidence favouring the OFS, first from the presence of Multimodal Object Files (subsection 7.1) and second from experimental research on artificial agents and non-human animals (subsection 7.2). Section 8 concludes and provides some suggestions for future approaches.

2. Preliminaries: specificities of MQ

Before delving into the core of the question, let´s take a closer look at the particularities of the thought experiment suggested by Molyneux more than three centuries ago.

First, MQ only applies to visuo-tactile interactions in shape perception. MQ presents a situation where a blind person observes, for the first time, stimuli only touched before. In this situation, the property involved is ´shape´. Shape properties are directly processed by sight, touch or both —no one can smell, taste or hear shapes. One can perhaps do it indirectly, e.g., when I hear a horn, I recognize a car, perhaps the shape of a car, but in no way this constitutes a direct perception. Other properties involve other senses: colour is exclusive to vision, temperature to touch, sweetness to taste, loudness to hearing and floralness to odour. Some perceptual properties are also perceptible across different senses, sometimes providing overlapping information for objects and scenes in the environment. Cross-modal perception appears very early in infants (Purpura et al., 2018; Joanne et al., 2014) and continues developing throughout our lives (Norman, 2006).

It would not make much sense to question whether a subject who suddenly recovers from congenital deafness would recognize his mother's voice, nor whether a subject who suddenly recovers from congenital blindness could visually recognize a rose previously smelled. The answer to these questions is clearly negative,[6] which shows that, at least intuitively, the case posed by MQ

---

[5] Although OFS says nothing about (1) and (3), it can still be adjusted to both by introducing minor modifications. After all, the only difference between (1), (2) and (3) is where we put the accent, whether on identity, referentiality or structure. Although (1) and (3) are, I think, harder to defend, I will not argue here against them; I simply assume that OFS is more consistent with (2).

[6] Cases of synesthesia are rare and will not be considered here.

reveals a particular conjunction between vision and touch in recognizing objects from shape properties. The specificity of shape for vision and touch is, therefore, essential to understanding MQ, which means that any response to MQ will not be, in principle, generalizable to the relationship between other sensory modalities. Note also that other properties like size are also perceptible through vision and touch, but size is not the type of property that produces the typical controversies that the MQ original case of shape produces. It seems evident that a newly sighted can visually differentiate between a basketball and a golf ball previously touched. Similarly, the difference in weight between the sphere and the cube (imagine the sphere weighs ten times as much as the cube) can only be appreciated by touch but not by sight. Consequently, much caution when extending the MQ case to any possible combination between other senses or different properties than shape.[7] MQ is, therefore, a very specific case.

Second, MQ only works in the appropriate direction. The reverse situation, i.e., if a person who can visually distinguish between a cube and a sphere would recognize it solely by touch without any prior tactile experience, seems less controversial.[8] Hardly one can replicate a case like this, but the reverse option offers quite a different impression. In this case, the affirmative response is much more applicable than the negative one. It can be argued that when visually perceived, the differences between stimuli are captured and registered with greater precision and strength than when perceived tactilely. Most likely, this is because visual recognition produces a more extensive imprint than tactile one. The reverse situation has already been explored in the field of robotics. Falco et al. (2019) showed that robots are able to recognize through tactile exploration previous visual exposition to objects without having touched any object before. Again, this is an unsurprising discovery since visual information is, by far, more informative than tactile one, at least regarding object shape recognition. However, the positive response to the reverse situation opens the possibility of deeply exploring a positive response to the original characterization. All this shows that the specificity of MQ requires examining the visuotactile multimodality in the direction suggested by the original case, but also that the mechanism by which visual shape information is transferred to the tactile one in the reverse situation can also work for the original case. Perhaps, the prevalence of vision over touch clouds the informational transfer from touch to vision, but not from vision to touch. In subsequent sections, this will be further explored.

Third, MQ involves a single content delivered in different formats. Indeed, it is easy to see that the same content can be manifested in various formats. For example, the content "that is a cube" can be represented in spoken English, written Spanish, illustrated in sign language, codified in Morse or even photographically. In principle, the same occurs in the opposite direction —different contents can be depicted in the same format. For example, the same artwork may have different content for an expert than for a non-expert, or the same word can have different contents, e.g., homonymous words. So, content and format are both parts of a representation, but they can be individuated.

7 This is crucial for the visual-centric bias in perception research (for interesting insights into why vision dominates perception research, see Hutmacher, 2019). The idea that everything we know about vision extends to other sensory modalities falls apart.

8 Surprisingly, researchers have not taken this option into account. I have only read this variant in Bruno and Mandelbaum (2010), who also show astonishment about the lack of attention to the reverse case (see footnote 27).

The question raised by Molyneux places us in the first situation. According to MQ, similarities in representation might imply similarities in representational content but not in representational format. That is, the same content can be depicted in different formats, visually (iconically) or tactile (sensitively), but just as the representational content of seeing a sphere and the representational content of touching a sphere are similar, the format in which such content is represented differs. Therefore, in the MQ case, visual and tactile representations are felt phenomenally different because they are conveyed in different representational formats, but coincide in their representational content.[9] So, any positive answer to MQ must focus their arguments on the similarities of representational content.

To sum up this section: 1) MQ is a specific case involved in visuotactile interactions not extended to other perceptual systems, 2) MQ only works in the direction offered by the original case (the blind person that suddenly sees) when we take the reverse direction (the person who previously couldn't feel touch and suddenly gains the ability to do so), it does not work, 3) any positive response to MQ should focus on the content rather than the format in which such content stores in memory.

Now, once paved the ground let me show how there can be a compelling interpretation of option 2.

3. Can we solve MQ empirically?

Arguably, MQ hardly yields a clean and non-controversial experimental situation. The main reason is that the sudden restoration of sight with no other concomitant issues may not be realistic. The effects of congenital blindness on the mechanisms of vision (optical or cognitive) require significant time for adaptation from sight restoration to object shape recognition.[10] This objection has been put in different ways by past philosophers (La Mettrie, 1750; Smith, 1795) and continues to be the most significant objection to empirically testing MQ (Degenaar, 1996; Meltzoff, 1993; Cattaneo and Vecchi, 2011).

Despite this, much of the recent interest in MQ is motivated by the possibility of testing it experimentally. Technical and experimental advances urge us to reconsider old questions with new tools, and MQ is one of them. There are relatively few studies of subjects recovering from congenital blindness, which undoubtedly is a symptom of the scarce experimental guarantees traditionally offered about MQ. However, studies of subjects recovering from cataracts stand as one of the most promising experimental situations to address MQ.

---

9 I am not denying that phenomenological similarity is important for the categorization and identification of objects themselves (there are strong empirical reasons to think like that) but in my view, the particularities of the MQ case require, for a positive answer, focusing on coincidences in representational content (for a more nuanced discussion see Di Stefano and Spence, 2024a).

10 Almost from its inception, philosophers have suggested modifications to the original formulation. Diderot (1749) suggested using two-dimensional shapes to avoid initial failures in the visual processing of tridimensionality, Reid (1764) recommended using geometricians as subjects, and Leibniz (1760) suggested offering the subject some clues about the type of forms that must identify. More recently, Levin (2008) suggests providing subjects enough time for complete visual restoration while cancelling their ability to identify shapes by sight alone. Be as it may, and in light of this concern, MQ remains as follows: If the Molyneux subject were able to make a sufficient 'visual discrimination' between shapes, would she be able to recognize which is the cube and which is the sphere?

An often-cited and much-discussed study conducted by Held et al. (2011) was thought, at least for a while, to have resolved the question towards a negative answer. These researchers examined five patients soon after cataract surgery for congenital blindness. The task involved two steps. In the first one participants were visually or tactilely presented with one target stimulus (Lego blocks). The second step was to identify from two further stimuli, again tactilely or visually, which one is identical to the target stimulus. Subjects' performance was compared across three conditions: the visual-visual condition, which of the seen shapes is identical to the seen target one; the tactual-tactual condition, which of the felt shapes is identical to the felt target one; and the tactual-visual condition, which of the seen shapes is identical to the target felt. The results indicated that in the two first conditions (visual-visual and tactual-tactual), subjects performed well —suggesting they were able to see and feel the objects— but their performance in the visual-tactile condition was very poor, approximately at chance. The experimenters concluded that the newly sighted subjects did not exhibit an immediate transfer of their tactile shape knowledge to the visual domain (p. 552), and, therefore, the answer to the Molyneux question is likely negative. However, and leaving aside important methodological concerns (see Connolly, 2013; Schwenkler, 2012, 2013; Cheng, 2015; Clarke, 2016; Levin, 2018), the experiment does not solve the pressing concern of the sudden restoration. In fact, experimenters observed that subjects could not visually recognize the stimuli within 48 hours after surgery, but they recognized them a few days after recovering their sight. Accordingly, the answer to MQ is negative immediately after sight restoration but turns positive after a short period.

In sum, the experimental literature discussing these issues has not offered, to date, a convincing response to the issues surrounding MQ. For many, the possibility of complying with the necessary conditions to run an experiment that could provide a persuasive response to MQ is remote (Cheng, 2015). Given all this, MQ appears as a thought experiment that, *at least in adult human subjects*, hardly can be transformed into a testable one. The experiment requires an abrupt recovery of the visual system, which is not very realistic in practice (although empirically possible). So, despite efforts to move the thought experiment into the realm of experimental practice, pending the definite experiment (if workable), any answer to MQ will ultimately require a certain amount of inference.

Alternatively, empirical research concerning MQ has also been developed in non-humans. It is worth highlighting, therefore, two lines of research that could be especially promising. The first comes from robotics (Falco et al., 2019; Liu et al., 2019), and the second from animal experimentation (Solvi et al., 2020; Versace et al., 2024). All these lines result in an affirmative answer to MQ and will be deeply addressed later.

4. *Evidence on cross-modal perception*

Much of the positive answers to MQ are based on recent developments in cross-modal object recognition. Cross-modal object recognition is defined in neuroscience and psychology as *the ability to recognize an object, previously inspected with one modality e.g., touch, via a second modality, e.g., vision, without prior training in the second modality.*[11] An initial and fundamental step towards a positive answer to MQ is understanding that the different sensory modalities usually interact, overlap, modulate, integrate or benefit each other to capture the most appropriate interpretation of perceptual

[11] "Psychology dictionary," http://psychologydictionary.org/.

representations. This is often called intermodal perception. Extensive evidence shows that interactions between the senses are robust, abundant, varied and occur at a very early stage of perceptual processing (O'Callaghan, 2015). Of course, there are perceptual properties whose representational content is exclusive of a particular sensory modality —e.g., there is no such thing as perceiving colours through other modalities than vision— but the interaction between two or more sensory modalities is a very common situation for the representation of the vast majority of object properties (such as shape properties, perceptible by both vision and touch). There are two types of inter-modal effects in perception: multimodal perception (or multimodal integration) and cross-modal perception. The former occurs when two or more sensory modalities fuse to detect or identify stimuli (e.g., audio-visual speech), and the latter when an object initially perceived by one sense modality is recognized by another sense modality. Clearly, MQ falls in the second type of effect.

The concurrence of vision and touch in shape perception is not controversial; however, a response to MQ goes beyond the cross-modality occurrence. Affirmative and negative answers can easily accommodate this evidence —MQ does not question whether there is informational transmission from touch to vision but whether the learned tactual information is visually available even without visual information ever entering the retina before. Generally, studies on cross-modal perception do not consider these effects in blind people since visuo-tactile cross-modality is available to normal sighted subjects but not to blind ones. That is, cross-modal recognition is amply demonstrated for subjects with intact vision and touch, but MQ requires that cross-modal recognition exists in the absence of previous cross-modal experience; the specificity of the MQ case forces us to reconsider this issue deeply.

One potential solution to this issue is to consider that perceptual information is translated into a common code (Altieri, 2015), i.e., that there is a shared space for both sensory modalities in which the information is commonly encoded. This seems to occur in crossmodal transfer between haptic and visual representations. Although haptic and visual channels are independent in the earlier levels of processing, at the later levels, their signals are commonly encoded (Held, 2009). The lateral occipital complex (LOC), a region of the ventral visual pathway, has been proposed as the brain area where visual-haptic signals concur. This typical visual area not only responds to haptic 3-D and 2-D stimuli (Amedi et al., 2002; Prather et al., 2004) but also processes shape information independently of the sensory modality used to acquire it (Lacey and Sathian, 2015).[12]

Now, what type of information is commonly encoded in crossmodal perception? Let´s consider two possibilities. The first is that the incoming bottom-up structural information of the stimuli transfers from one sense to another. The second is that previously stored top-down conceptual knowledge transfers from one sense to another. The first option is not very realistic since vision and touch use different sense channels to process the structural information of objects. Recall that

[12] The most recent studies in neuroscience show interactions between vision and touch in the cerebral cortex (Sathian and Lacey, 2022). These crossmodal interactions occur, for example, when haptic stimuli activate visual brain regions (Snow et al., 2014), thus converging visual and haptic stimuli. These convergences occur in the association brain areas, mostly in the Lateral Occipital Cortex (LOC). These typically vision-specific areas are activated in the absence of retinal input (Amedi et al., 2001) and encode haptically perceived shapes even for congenitally blind subjects (Amedi et al., 2010). Moreover, LOC plays a key role in both visual and tactual object recognition (James et al., 2006), suggesting that it integrates the information processed in both modalities.

vision and touch share informational units in the later levels but not in the earlier ones (Held, 2009). The second is simply unsustainable since in the absence of incoming information we cannot speak of perception properly. In my view, cross-modal perception requires the conjunction of both types of transmission: crossmodal transference of some structural properties of objects and crossmodal transference of the previously stored knowledge. Let me illustrate it better. The tactile system uses a generic structural shape property (e.g., the typical roundness of spheres) to ground the perceptual identification (i.e., sphere), and fundamentally, this is achieved by way of perceptual reference. Referential information acts, therefore, as the conjunction of the bottom-up incoming information and the top-down stored one. This is what is transferred from one sense to another, regardless of which one of the two types of encodings (bottom-up or top-down) acquires more weight in stimuli identification, the central point is that perceptual reference encloses the two encodings into a single one. Therefore, if we understand referentiality as the causal relation between referents-in-the-world and mental representations that refer to them, then becomes reasonable considering that it is the confluence between both types of informational units (i.e., referentiality) what it is transferred from one sense to another. In sum, when two mental representations co-refer, they represent the same object or property (Rescorla, 2020).

Let´s put all this in MQ terms. Perceptual reference requires both the ability to perceptually distinguish between objects (cube and sphere) and the ability to recognize or identify each one of them. On the one hand, subjects can differentiate between objects just by responding merely to differences in the incoming low-level spatial features without involving identification —subjects can find differences between the two objects without knowing which is which. On the other hand, subjects cannot identify stimuli whose properties do not arrive at their senses —perceptual identification cannot float free from low-level structural features. Thus, only if the informational exchange between tactile and visual perceptual systems is referential the newly sighted subject will be able to recognize crossmodally tactilely familiar shapes by sight alone. In short, subjects learn to differentiate tactilely between the sphere and the cube by creating a referential correspondence between the salient properties of each (e.g., cube with squareness and sphere with roundness) with the correspondent identificative label attached, thus referentially anchoring the cube and the sphere with their salient properties. Now, this referential anchorage should be registered through learning in order to be subsequently recovered through memory and, therefore, recognized, i.e. there will not be recognition without previous referential anchorage.

Suppose this story is correct. Suppose the MQ subject visually distinguishes between the sphere and the cube because her brain referentially matches the top-down conceptual information previously registered by touch with the bottom-up spatial information incoming by sight. In sum, subjects are able to crossmodally recognize objects that have never crossmodally experienced. Now, by what mechanism can this occur? The next section suggests that *object files* can serve as this mechanism.

5. Object Files as the mechanism for cross-modal perceptual recognition

This section introduces how referential informational content is stored in object files for subsequent recovery when required. Treisman et al. (1983) first defined object files as "the temporary representation in which the information that pertains to a particular object accumulates and is updated when the object changes" (p. 531). Since then, the notion has been widely employed in many and varied areas of the psychology and philosophy of perception. Despite a considerable lack of consensus on how object files should be understood, Green and Quilty-Dunn (2021) have recently deployed a thorough characterization.[13] According to these authors, an object file is a perceptual object representation generally characterized as a representation that (i) sustains reference to an external object over time and (ii) stores and updates information concerning the properties of that object (p. 666). The importance of the first point is that the storage of the information is referential and sustained in time. The second point suggests that object files are dynamic and may change over time if required by environmental situations. To this general characterization, they add two important points: first, object files represent information in a propositional format, and second, object files store information concerning different feature categories in separate memory stores (p. 666). According to the first point, differences in format between visual and tactile representations (iconic and tactile formats) are irrelevant for subsequent recovery since object files employ a discursive propositional format (i.e., a minimal sufficiently abstracted way to store information). The second point not only assumes that object files are, in essence, memory mechanisms but also that these mechanisms store features of different categories into separate memory stores within an object file.

According to Green and Quilty-Dunn (2021), object files are a construct empirically supported by experimental research. They appeal to three different experimental paradigms. The first are those experiments that examine our ability to maintain and update representations of objects as they move. For example, in the Multiple Object-Tracking (MOT) paradigm, participants should track a subset of flashing items from a bigger set of moving items. Participants are typically effective at tracking up to about 4 items (Pylyshyn 2003) despite significant changes in colour, shape, and size (Zhou et al. 2010) or despite intersecting each other or occluding by hidden barriers (Scholl and Pylyshyn, 1999). MOT suggests that there must be a mechanism —object files— that encodes spatiotemporal addresses of objects to maintain identity over time and across changes (Quilty-Dunn, 2020).

The second set of experiments are those that examine our ability to maintain representations of objects after they have disappeared from view; this is the Object-Specific Preview Benefit paradigm (OSPB). In OSPB, subjects are presented with two simple objects with a particular feature for each object (e.g., a letter). At some point, features disappear, and the objects move to new locations. Then, one of the features appears in one object, and subjects should indicate whether it matches one of the original features. Subjects´ performance is faster when the feature re-appears on the same object (compared to when it does on a different object). Kahneman et al. (1992) appeal to Object Files to explain this effect. The explanation is that responses are quicker because when the correct feature appears in the appropriate object, subjects benefit from the previewed information, which has been temporarily stored as a stable representation in an object file.

The third is a mechanism called Transsaccadic memory. Transsaccadic memory is the process by which remembered object information is updated across saccades (Irwin & Gordon, 1998; Hollingworth & Henderson, 1998). In order to build a coherent internal representation, the system should represent and update each saccade in memory; i.e., visual constancy, or the experience of

[13] For very recent criticism about the usability of this notion, see Block (2023) and Nanay (2023). I do not take part in these debates here; I simply assume that Object Files are a useful psychological construct that consistently characterizes how objects are perceptually represented, stored and updated.

the visual scene as continuous, is possible because internal representations are preserved across saccades recorded in visual memory (Higgins & Rayner, 2015). Under this paradigm, subjects are usually instructed to move their eyes toward moving objects and discriminate whether objects change their features or remain unchanged. The studies show that memory-based gaze correction is accurate, fast, automatic, and largely unconscious (Hollingworth et al., 2008; Irwin, 1992; Pollatsek et al., 1984), thus suggesting that representations stored in transsaccadic memory maintain phenomenal coherence across saccades. It is argued that object files modulate this process by maintaining a representation of objects across changes in retinal stimulation caused by eye-movements, i.e., object files enable us to keep the object's identity across saccadic movements (Gordon and Vollmer, 2010; Schut et al. 2017; Quilty-Dunn and Green, 2022). Gordon and Vollmer (2010), for example, found a decrease in reaction time for naming a previewed object. This object-specific effect was significantly reduced when typical properties of objects changed, but not when changes did not involve the typical properties of objects —e.g., the correspondence between transsaccadic representations of a banana depending on the yellow colour of the banana, whereas correspondence for transsaccadic representations of an object without a characteristic colour (such as a cube) does not depend on its colour (see also Gordon, 2014). This shows that transsaccadic memory holds the representation from low-level properties and their relation to abstract categories, thus suggesting that abstract information is referentially bonded into object files for subsequent recovery.[14]

Object files act, therefore, as a database that amodally preserves referential perceptual information over time and consists of category-specific slots propositionally stored in memory for subsequent recovery.

6. *Object files and MQ*

The question at hand is, therefore, how object files can help determine a positive answer to MQ. Specifically, we want to know if object files are responsible for recognizing information that has never been seen before. Importantly, it would be inaccurate to say that, according to this approach, the information is literally transferred between sensorial modalities. Instead, the approach suggests that the information is available for any sensorial system capable of interpreting it —there is not, properly speaking, a transference of information from touch to vision, but a visual recognition of referential information previously acquired from touch or any other perceptual system. I argue, then, that object files can be thought of as the mechanisms responsible for storing referential information (regardless of the mode of acquisition), to be recovered in the future, by whatever mode of presentation compatible with the mode of acquisition. Of course, the more presentations, the stronger the imprint, and the easier it will be to recover. The reasons why this appears as a possibility are the properties usually ascribed to object files, properties that are mutually bonded to each other. I focus on four basic properties.

6.1 OBJECT FILES ARE REFERENTIAL

This is one of the main characteristics of object files, namely sustaining reference over time (Green and Quilty-Dunn, 2022). As stated above, by referential I understand the connection between the

[14] For more on the experimental paradigms that suggest the existence of object files, see Quilty-Dunn (2020) and Green and Quilty-Dunn (2021).

information that impinges the retina or the somatosensorial receptors with a specific conceptual knowledge acquired through learning. There is not much controversy on this point —the above-reviewed literature on the experimental paradigms base their results on the assumption that object files are referential. One way to see this straightforwardly is by considering the difference between perceptual reference and perceptual attribution. We have just seen that referential content may survive some modifications of low-level properties of objects—phenomena like amodal completion (perceptual object recognition even with extensive occluded parts of the objects) or experimental research on multiple object tracking (perceptual object recognition even with changes in low-level features of objects) do not admit discussion on these effects. It seems then that perceptual representation comprises at least two aspects: reference to perceptible individuals and attribution of properties to those individuals. The paradigms described above show that object files are able to maintain the identity of objects even with changes in spatiotemporal, shape, or colour properties. Although these properties are potentially relevant to determining perceptual reference to the object, they are not applicable by virtue of being encoded in the perceptual object representation for that object. Thus, object files are, in this crucial sense, referential.[15]

6.2 OBJECT FILES ARE AMODAL

From the previous property, one can directly uphold that object files must store object representations amodally.[16] Examples of amodal representations are feature-based word meaning representations, semantic networks, schemata, or frames of reference. This is central to sustaining the affirmative answer suggested by OFS. MQ subjects not only tactilely perceive the roundness of the sphere or the cube's edges but they also learn to differentiate them conceptually. In this sense, the tactile recognition of the stimuli involves both the contact of the somatosensorial receptors with an external object as well as the referential identification of the object as a distinctive exemplar of a particular category. Once physical contact occurs, the stored referents corresponding to the recognition of a given stimulus connect with the incoming information. Therefore, the internal amodal representations may capture information from one or more modalities, but these representations are, themselves, modality-unspecific, i.e., they constitute an abstract description of what they represent. Object files are, in this crucial sense, modality-independent internal representations of objects.[17] Evidence for this internal amodal object representation has been found in humans (Jordan et al., 2010), mammals (Winters and Reid, 2010), fish (Schumacher et al., 2016) and even in invertebrates such as bees (Solvi et al., 2020). Now, although the amodal structure is manifestly different from the physical structure of the things represented, object files

[15] There is a very interesting debate on whether perceptual attribution determines or guides perceptual reference (for discussion, see Quilty-Dunn and Green, 2023). Some think that all attribution guides reference (Burge, 2010), others that none attribution guides reference (Pylyshyn, 2009) and others that only a privileged subset of attributions guide reference (Flombaum et al., 2009). This is not relevant for the proposes of this paper, but there are reasons to think that the range of perceptual attributives used to guide reference shifts adaptively with context (Quilty-Dunn and Green, 2023).

[16] For a very recent review of the different uses of the amodal sensory dimension, see Spence and Di Stefano (2024a).

[17] Spence and Di Stefano (2024a, 2024b) have raised doubts on whether shape information is coded amodally by the senses of vision and touch. In the absence of convincing empirical evidence to support the claim of amodal sensory qualities, they suggest restricting the term amodal to abstract cognition rather than sensory perception. This might, in principle, contradict my view, but I think compatibility is possible. Simply, my understanding of object files as a cognitive mechanism that stores amodal and abstract information that mediates, guides and configurates sensory perception, is not at odds with the assertion that shape information is initially coded in a modal way. I am grateful to an anonymous reviewer for bringing this literature to my attention.

allow the correspondence between external and internal objects; correspondence that can be decoupled from specific sensory modalities. In sum, once the low-level features of an object are integrated, the system stores via object files an abstract amodal representation of this object, which will be available for subsequent recovery; i.e., once amodal storage is supported, the incoming information in one modality may be recognized through another. Therefore, amodality becomes modal at the very moment the senses intervene. That is, when we see an object we have never seen before but about which we have previously obtained information through touch, the amodal information stored with the representational content of that object becomes modal at the very moment that vision intervenes. If this is so, then MQ would move towards a positive response.

6.3 OBJECT FILES ARE PROPOSITIONAL IN FORMAT

In strong relation to the previous points, object files store information in a propositional format rather than in an iconic or tactual one. In the preliminaries of this paper, I sustained that any positive answer to MQ must come from the content of the perceptual experience rather than the format in which such content is represented. I think this is clear when we go straight to the experience: if a response to MQ depends on the format in which the information is recorded, then visual and tactual information are depicted in different formats and, therefore, a positive answer to MQ is not possible. For this reason, I argued that any positive answer must involve the content represented. However, the format in which objects are recorded in object files is independent of the format in which it is perceived —objects can be perceived iconically and recorded discursively. Indeed, the referential character of object files and their amodality indicates that they represent their contents in a non-iconic propositional format. This is key for subsequent recovery of the informational content independently of the format in which such content is perceived. Let me expand on this. Formats are general structural properties of representational states that play a role in individuating types of representations. According to Kosslyn et al. (2006, p. 8), 'A format is a type of code'. Just as photographs are codified in an iconic format or tactile graphics are codified in a tactile format, sentences are codified in a discursive or propositional one. For example, the sentence 'that is a rose' is structured differently from a picture or a tactile graphic of a rose. All of them share content (rose) but differ in the format in which such content is represented (discursively, visually or tactilely). There are, however, crucial differences between the properties represented in one or the other. Just as perceptual-like representational formats (visual and tactile in this case) can represent low-level properties such as shape, colour or orientation, the properties represented in discursive formats are silent to this information (at least in this case). The discursive representation, on the other hand, provides categorical information (rose) that a layperson in flowers cannot grasp from the perceptual-like representational formats. The idea is that object files are codified discursively, independently of the mode in which the information has been presented (see Quilty-Dunn, 2016). All this is in strong concordance with the positive answer to MQ defended here and straightforwardly connects with the following property of object files.

6.4 OBJECT FILES ARE CATEGORY-SPECIFIC

We have just seen that object files retain referential information abstractly, propositionally and independently of the modality in which the information has been recorded. A direct consequence is that object files behave as category-specific previews used to determine the referential matching with external objects. A preview benefit, for example, requires that matching feature combinations are ascribed to a specific object category; although for features not initially attributed to the reviewed object category, there will be no benefit (O´Callahan, 2014, p. 11). The idea here is that even with changes in some object properties the system recognizes objects as pertaining to specific categories, i.e., a handful of properties can be diagnostic of the objects category. Green and Quilty-Dunn (2021) argue that feature representations are organized into separate category-specific slots located within an object file. When subjects are asked if they recognize objects (e.g., a sphere), they represent representations of individuals that function computationally as pointers that enable access to these category-specific slots. The case is that this type of storage is not specific to any sense modality. Subjects can store the category 'sphere' with its typical properties into an object file by touch, and when subsequently asked if they are visually facing a sphere they will retrieve the correspondent object file that stores the individual categories of the conceptual references (e.g., sphere) and the feature categories of the spatial properties of spheres (e.g., roundness). According to Quilty-Dunn and Green (2021), this is corroborated by evidence that indicates that "features from the same category compete with one another to a much greater degree than features from different categories, even when the features are bound to the same object"(p. 685). One outstanding consequence of this type of storage is the extremely fast time in which humans categorize objects. Potter et al., (2013), for example, found object recognition as fast as 13ms. from stimulus onset. This extremely fast recognition is a substantial adaptative quality since it allows object recognition without the cost of perceptually codifying too many low-level properties; perhaps only a salient feature is enough for object recognition. This property of object files allows object recognition by retrieving the salient properties stored into the object file that administrates the specific category; again, object files are not modality-specific and therefore nothing prevents this process can being done from whichever sense modality is able to recognize it.

6.5 TAKING STOCK

Once the properties of object files are defined, it is easy to see the mechanism from which a newly sighted subject might recognize tactilely familiar shapes by sight alone. If tactile information is stored in object files in the way defined above, i.e., referentially, amodally, discursively and category-specifically, then this information could be subsequently retrievable throughout the different sense modalities able to decode the type of information deployed. Since shape perception can only be registered visually or tactilely, it will be retrieved from the correspondent object file in one of the two modalities. Thus understood, appealing to object files can offer a positive answer to MQ.

7. Defending OFS

Once OFS is explained, I will provide some empirical research that involves the view. One is the possibility that object files store multimodal information, the other concerns experimental research in robotics and animal experimentation. The former shows that object files store information in a way that transcends sensory modalities, whereas the latter opens up the possibility of connecting empirical research on crossmodal perception with object files, thus providing an avenue towards

OFS. Although neither of these is definitive, they both constitute a step towards the OFS's positive response to MQ.

7.1 MULTIMODAL OBJECT FILES

Jordan et al. (2010) conducted a study to determine whether the contents of object files are specific to a particular sense (modality-specific) or if they can be accessed across different senses (amodal). The researchers carried out two experiments. The first experiment investigates whether object file correspondence can occur through a combination of visual and auditory information. The second experiment rules out the possibility of explaining the results in the first experiment via verbal encoding. The study indicates that object files are not strictly tied to vision and that object-related information is stored in an amodal format. Therefore, an object file can be initially formed with visual input and later accessed with corresponding auditory information, indicating that object files may operate at a multimodal level of perceptual processing. Researchers conclude that object files "store object-related information in an amodal format that can be flexibly accessed across senses" (p. 500). They leave open the possibility that object files are modality-specific but accessed by any modality able to recover the information, i.e., that there is a cross-modal correspondence, or that object files are devices that store information amodally, in a sort of common code mechanism, in order to be accessed later by any modality whatsoever. Any of these possibilities lean towards an affirmative response to MQ, although the second one is more aligned with the objectives of the OFS.

7.2 ROBOTICS AND ANIMAL EXPERIMENTATION

As I pointed out above, experimental research on MQ with blindness restoration in humans is inconclusive due to the implausibility of sudden restoration of the visual system. However, alternative approaches involving robotics and non-human animals might provide valuable insights. Studies on robotics (Falco et al., 2019; Liu et al., 2019) and animal experimentation (Versace et al., 2024; Solvi et al., 2020) offer compelling evidence that aligns well with the OFS framework.

Falco et al. (2019) trained a classifier by using visual data stored from a camera to recognize objects only with tactile data, without any prior tactile information. This approach represents the opposite scenario from MQ, and as previously mentioned, it is less contentious than the original one. As said, this is likely because of the dominance of the human visual experience over the tactile one, but this dominance can be circumvented by building an artificial system. The point is that the very same mechanism can likely be deployed for both visual-to-tactile and tactile-to-visual transmission of information in humans. These researchers stem from the idea that crossmodal perception necessarily requires common information between the modalities. The way they find this common space is through descriptors. A descriptor is an abstract symbol (words or characteristic features) that collects the attributes of items and stores them in databases to posteriorly enable the system to recognize objects based on data collected (see also Zhang et al., 2016). In this way, the system can recognize objects in any sense modality by matching the information entered from sensors with the descriptors stored in databases. Indeed, although these researchers do not directly refer to an object file mechanism, something similar can be inferred.

Likewise, other researchers interested in developing a mechanism to implement an artificial active visual-tactile cross-modal matching designed a shared dictionary learning model which can simultaneously learn from the information coming from the visual and tactile sensors (Liu et al., 2019). All of this is reminiscent in many ways of the referential, amodal, propositional and category-based mechanism proposed here. The crucial point is that to design an artificial mechanism capable of transferring knowledge across various perceptual modalities, researchers must move beyond sensory modalities and leverage algorithmic abstract information.[18]

Of course, one can say that the mechanism generated for object recognition in artificial agents might be far from mirroring the mechanism naturally employed by humans, but in fact, mirroring human behaviour is precisely what researchers in robotics try to do. I have no space here to elaborate further on this issue, but I think that studies in robotics can offer important clues about MQ, mainly because it easily manages to overcome the insurmountable limitations of human experimental situations. Be as it may, it seems that this line of evidence is close to what is suggested in the present paper.

The other line of evidence favouring an affirmative response to MQ consistent with OFS comes from non-human animal experiments. For instance, Versace et al. (2024) recently conducted a experiment with chicks. They exposed newly hatched chicks in darkness to either tactile smooth or tactile bumpy stimuli for 24 h, and immediately after they tested them in a visual recognition task. During their first experience with light, chicks exposed to smooth tactile stimuli approached the visual smooth stimulus significantly more than those exposed to bumpy tactile stimuli. It is noteworthy that chicks are precocial; i.e., they hatch with mature visual, proprioceptive and motor systems, so that their perceptive and motor responses can be evaluated already in the first hours of life, thereby avoiding the rapid restoration problem encountered in empirical research involving humans. These findings show that visually inexperienced chicks can solve MQ positively, thus indicating that cross-modal recognition does not require previous multimodal experience. They argue:

> This significant difference in visual preferences in chicks that differed only in tactile experience shows that visually naive chicks learn about objects experienced in the solely tactile modality, and can use representations based on tactile experience to solve a visual recognition task. The ability of newly-hatched chicks to discriminate between visual objects at first sight, based on previous tactile experience, solves Molyneux's problem, showing that cross-modal recognition from tactile to visual sensory modality does not require previous experience with simultaneous multi-modal stimuli. (Versace et al., 2024, p. 4)

The key point is that chicks are capable of spontaneous cross-modal recognition in the absence of previous cross-modal experience. While not explicitly relying on an object file system, the authors predict the existence of an innate system for crossmodal shape recognition. But crucially, this crossmodal system can only be active when chicks first incorporate tactile information (cross-modality is not possible if no stimulus has ever been experienced). This opens the possibility of abstractly storing tactile information into an object file for later visual retrieval. If chicks are indeed born with an innate crossmodal shape system mediated by object files, it is likely to be quite

[18] For a recent overview of visuo-haptic object perception for robots, see Navarro-Guerrero et al. (2023).

rudimentary but, despite all, such a system enables them to store amodal internal representations of object shape properties.

A similar experiment was conducted to investigate crossmodal perception in bumble bees. Solvi et al. (2020) trained bumble bees in the dark to discriminate between spheres and cubes to posteriorly test their visual discriminative ability for the same objects that could only be seen (in the light) through a clear barrier, and not touched. The bees showed a preference for objects that had been previously rewarded in the dark. Conversely, bees trained to visually discriminate the objects spent more time with the objects in the dark (tactile) setting. Researchers conclude their findings as follows:

> Whether bumble bees solve the task by storing internal representations of entire object shapes (cube or sphere) or local object features (curved or flat edge) remains unknown. In either case, our experiments show that bumble bees are capable of recognizing objects across modalities, even though the received sensory inputs are temporally and physically distinct. Bumble bees show a kind of information integration that requires a modality-independent internal representation. This suggests that similar to humans and other large-brained animals, insects integrate information from multiple senses into a complete, globally accessible, gestalt perception of the world around them. (p. 911)

Again, the connection with OFS is clear. Bumble bees retain modality-independent (amodal) internal representations that can be accessed by any perceptual modality capable of decoding them.

This body of evidence supports the OFS approach suggested here. In addition to this, other cross-modal recognition studies in animal cognition, such as cetaceans (Bruce and Pack, 2022) or rats (Reid et al., 2014), as well as cross-modal recognition studies in newborns (Streri, 2012) suggest that the mechanism through which perceptual information is stored and remains accessible for subsequent recovery is an object file mechanism (i.e., a referential, amodal, propositional and category-specific mechanism). It is evident that the complexity of the object file mechanism in adult humans far exceeds the simpler mechanisms found in bees, chicks, and other animals —the rich conceptual repertoire of humans is unparalleled in the animal kingdom. Be that as it may, these studies along with previous research reviewed on robotics, can shed light on the mechanism underlying crossmodal perception and consequently offer valuable clues on how to effectively respond to MQ.

8. Conclusion and future approaches

In this paper, I have argued for a positive answer to MQ based on the Object File Strategy (OFS). Firstly, I discuss the specificities of MQ and argue against its experimental assessment. Then, I focus on crossmodal recognition studies, which demonstrate the interaction between vision and touch in shape perception. However, MQ suggests a different scenario. Despite the inequivocal interaction between the two sensory modalities, the defenders of a negative answer to MQ can claim that if subjects have never experienced visual stimuli, then visuo-tactile cross-modality is not possible. Experiments on crossmodal recognition, indeed, do not consider these effects in blind people.

Once here, I have argued that appealing to OFS can provide insight into these issues. The way and the type of information encoded in object files are crucial to my argument. I argue that if tactile information is encoded referentially and the information encoded is amodal, propositional and category-specific, then the information encoded in object files will be available for any sense modality capable of decoding it and, therefore, a newly sighted might recognize tactilely familiar shapes by sight alone. To support this, I have refered to empirical research. Firstly, I draw on Jordan et al. (2010) who demonstrate that object files operate at a multimodal level of perceptual processing, suggesting that object-related information should be stored in an amodal format and therefore it will be flexibly accessed across senses. Secondly, I cite studies on crossmodal recognition in artificial agents and non-human animals, which usually appeal to some sort of common code mechanism. This research not only avoids the experimental pitfalls of human research but also aligns perfectly with OFS.

If all this is on the right track, then an avenue to reconsider a positive answer to MQ opens before our eyes. I urge researchers to delve into MQ by considering the way in which perceptual information is received, stored and retrieved. Since the notion of object file inexorably appears as a crucial element throughout this process, I encourage those interested in this topic to consider these issues from the perspective of OFS.

Finally, I suggest researchers consider OFS through other theories on mind functioning. For instance, OFS is in line with the postulates defended by the Predictive Coding Framework according to which minds predict sensorial information by updating the global world model inserted into our brains (Clark, 2013), and fits perfectly with the Language of Thought hypothesis, the idea that thought is done in a mental language. The Language of Thought hypothesis, first suggested by Fodor (1975), has been currently retrieved by Quilty-Dunn, Porot and Mandelbaum (2022) by extending it to perception. I strongly suggest researchers take these lines of coincidence to put together a mutually consistent theory that englobe all these elements and tries to deploy all of them in a single hypothesis. I prognosticate that such a theory inescapably would offer a positive response to MQ in the terms suggested by the present paper.

Aknowledgements

I am grateful to the Research Group in Analytic Philosophy (Appearances, Singular Content and Fiction), of the Logos Group at the University of Barcelona, for the partial funds they provided me for presenting the results of this research project in the "XI Congreso de la Sociedad de Lógica, Metodología y Filosofía de la Ciencia en España".

References

Altieri, N. (2015). Multimodal theories of recognition and their relation to Molyneux's question. *Frontiers in Psychology, 5,* 1547. https://doi.org/10.3389/fpsyg.2014.01547

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. https://doi.org/10.1038/85201

Amedi, A., Jacobson, G., Hendler, T., Malach, R., & Zohary, E. (2002). Convergence of visual and tactile shape processing in the human lateral occipital complex. *Cerebral Cortex, 12*(11), 1202–1212. https://doi.org/10.1093/cercor/12.11.1202

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. https://doi.org/10.3233/RNN-2010-0503

Berkeley, G. (1709/2008). An essay towards a new theory of vision (NTV). In D. Clarke (Ed.), *Philosophical writings* (pp. 1–66). Cambridge University Press. https://doi.org/10.1017/CBO9780511802577.005

Block, N. (2023). Let's get rid of the concept of an object file. In B. McLaughlin & J. Cohen (Eds.), *Contemporary debates in philosophy of mind* (Chap. 27). Wiley-Blackwell. https://doi.org/10.1002/9781394259847.ch27

Bruck, J. N., & Pack, A. A. (2022). Understanding across the senses: Cross-modal studies of cognition in cetaceans. *Animal Cognition, 25*(5), 1059–1075. https://doi.org/10.1007/s10071-022-01684-8

Burge, T. (2010). *Origins of objectivity*. Oxford, England: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199581405.001.0001

Campbell, J. (1996). Molyneux's question. *Philosophical Issues, 7,* 301–318. https://doi.org/10.2307/1522914

Cattaneo, Z., & Vecchi, T. (2011). *Blind vision: The neuroscience of visual impairment*. The MIT Press. https://doi.org/10.7551/mitpress/9780262015035.001.0001

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*. Routledge. https://doi.org/10.4324/9781315146935

Cheselden, W. (1728). An account of some observations made by a young gentleman, who was born blind, or lost his sight so early, that he had no remembrance of ever having seen, and was couch'd between 13 and 14 years of age. *Philosophical Transactions, 402*, 447–450. https://doi.org/10.1098/rstl.1727.0038

Clarke, S. (2015). Investigating what felt shapes look like. *i-Perception, 7*(1), 1–6. https://doi.org/10.1177/2041669515627948

Condillac, E. B. (1754/1930). *Condillac's treatise on the sensations* (G. Carr, Trans.). Favil Press.

Connolly, K. (2013). How to test Molyneux's question empirically. *i-Perception, 4*, 508–510. https://doi.org/10.1068/i0623jc

Degenaar, M. (1996). *Molyneux's problem: Three centuries of discussion on the perception of forms*. Springer. https://doi.org/10.1007/978-0-585-28424-8

Di Stefano, N., & Spence, C. (2024). Perceptual similarity: Insights from crossmodal correspondences. *Review of Philosophy and Psychology, 15*, 997–1026. https://doi.org/10.1007/s13164-023-00692-y

Diderot, D. (1749/1916). Letter on the blind for the use of those who see. In *Diderot's early philosophical works* (pp. 68–141). The Open Court Publishing Company. Retrieved from https://archive.org/details/earlyphilosophic00dide

Evans, G. (1985). Molyneux's question. In A. Noë & E. Thompson (Eds.), *Vision and mind* (pp. 319–349). MIT Press. https://doi.org/10.7551/mitpress/7111.003.0016

Falco, P., Lu, S., Natale, C., et al. (2019). A transfer learning approach to cross-modal object recognition: From visual observation to robotic haptic exploration. *IEEE Transactions on Robotics, 35*(4), 987–998. https://doi.org/10.1109/TRO.2019.2914772

Ferretti, G., & Glenney, B. (2021). *Molyneux's question and the history of philosophy*. Routledge. https://doi.org/10.4324/9780429020377

Fine, I., Wade, A. R., Brewer, A. A., May, M. G., Goodman, D. F., Boynton, G. M., Wandell, B. A., & Macleod, D. I. (2003). Long-term deprivation affects visual perception and cortex. *Nature Neuroscience, 6*(9), 915–916. https://doi.org/10.1038/nn1102

Flombaum, J. I., Scholl, B. J., & Santos, L. R. (2009). Spatiotemporal priority as a fundamental principle of object persistence. In B. M. Hood & L. R. Santos (Eds.), *The origins of object knowledge* (pp. 135–164). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199216895.003.0006

Fodor, J.A. (1975). *The Language of Thought*. Harvard University Press.

Gordon, R. D. (2014). Saccade latency reveals episodic representation of object color. *Attention, Perception, & Psychophysics, 76*(6), 1765–1777. https://doi.org/10.3758/s13414-014-0683-6

Gordon, R. D., & Vollmer, S. D. (2010). Episodic representation of diagnostic and non-diagnostic object color. *Visual Cognition, 18*(5), 728–750. https://doi.org/10.1080/13506280903004190

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., & Quilty-Dunn, J. (2021). What is an object file? *British Journal for the Philosophy of Science, 72*(3). https://doi.org/10.1093/bjps/axx055

Gregory, R. L., & Wallace, J. G. (1963). Recovery from early blindness. In *Experimental Psychology Society Monograph* (Vol. II, pp. 65–129). Heffers. https://digitalcommons.rockefeller.edu/jason-brown-library/92

Held, R. (2009). Visual-haptic mapping and the origin of crossmodal identity. *Optometry & Vision Science, 86*, 595–598. https://doi.org/10.1097/OPX.0b013e3181a72999

Held, R., Ostrovsky, Y., de Gelder, B., Gandhi, T., Ganesh, S., Mathur, U., & Sinha, P. (2011). The newly sighted fail to match seen shape with felt. *Nature Neuroscience, 14*, 551–553. https://doi.org/10.1038/nn.2795

Higgins, E., & Rayner, K. (2015). Transsaccadic processing: Stability, integration, and the potential role of remapping. *Attention, Perception, & Psychophysics, 77*(1), 3–27. https://doi.org/10.3758/s13414-014-0751-y

Hollingworth, A., Richard, A. M., & Luck, S. J. (2008). Understanding the function of visual short-term memory: Transsaccadic memory, object correspondence, and gaze correction. *Journal of Experimental Psychology: General, 137*(1), 163–181. https://doi.org/10.1037/0096-3445.137.1.163

Hutmacher, F. (2019). Why is there so much more research on vision than on any other sensory modality? *Frontiers in Psychology, 10*, 2246. https://doi.org/10.3389/fpsyg.2019.02246

Irwin, D. E., & Gordon, R. D. (1998). Eye movements, attention, and trans-saccadic memory. *Visual Cognition, 5*(1–2), 127–155. https://doi.org/10.1080/713756783

James, T. W., James, K. H., Humphrey, G. K., & Goodale, M. A. (2006). Do visual and tactile object representations share the same neural substrate? In M. A. Heller & S. Ballesteros (Eds.), *Touch and blindness: Psychology and neuroscience* (pp. 139–155). Lawrence Erlbaum Associates. https://doi.org/10.4324/9781410615671

Jao, R. J., James, T. W., & James, K. H. (2014). Multisensory convergence of visual and haptic object preference across development. *Neuropsychologia, 56*, 381–392. https://doi.org/10.1016/j.neuropsychologia.2014.02.009

Jordan, K. E., Clark, K., & Mitroff, S. R. (2010). See an object, hear an object file: Object correspondence transcends sensory modality. *Visual Cognition, 18*(4), 492–503. https://doi.org/10.1080/13506280903338911

Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology, 24*(2), 175–219. https://doi.org/10.1016/0010-0285(92)90007-O

Kosslyn, S. M., Thompson, W. L., & Ganis, G. (2006). *The case for mental imagery*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195179088.001.0001

Lacey, S., & Sathian, K. (2015). Crossmodal and multisensory interactions between vision and touch. *Scholarpedia, 10*(9), 795. https://doi.org/10.1016/B978-0-12-805408-6.00392-4

La Mettrie, J. (1996). Treatise on the soul. In A. Thomson (Ed.), *La Mettrie: Machine man and other writings* (pp. 41–73). Cambridge University Press. https://doi.org/10.1017/CBO9781139166713.006

Leibniz, G. (1981). *New essays on human understanding* (P. Remnant & J. Bennett, Eds. & Trans.). Cambridge University Press. https://doi.org/10.1017/CBO9781139166874

Levin, J. (2008). Molyneux's question and the individuation of perceptual concepts. *Philosophical Studies, 139*(1), 1–28. https://doi.org/10.1007/s11098-007-9072-5

Levin, J. (2018). Molyneux's question and the amodality of spatial experience. *Inquiry, 61*(5–6), 590–610. https://doi.org/10.1080/0020174X.2017.1372306

Liu, H., Wang, F., Sun, F., et al. (2019). Active visual-tactile cross-modal matching. *IEEE Transactions on Cognitive and Developmental Systems, 11*(2), 176–187. https://doi.org/10.1109/TCDS.2018.2819826

Locke, J. (1979). *An essay concerning human understanding* (P. H. Nidditch, Ed.). Clarendon Press. https://doi.org/10.1093/oseo/instance.00018020

Lotze, H. (1887). *Metaphysic* (B. Bosanquet, Ed.). Clarendon Press.

Martin, M. G. F. (1993). Sense modalities and spatial properties. In N. M. Eilan, R. McCarty, & B. Brewer (Eds.), *Spatial representations* (pp. 206–218). Basil Blackwell Ltd. https://doi.org/10.1093/oso/9780198238874.003.0012

Matthen, M., & Cohen, J. (2020). Many Molyneux questions. *Australasian Journal of Philosophy, 98*(1), 47–63. https://doi.org/10.1080/00048402.2019.1603246

Meltzoff, A. N. (1993). Molyneux's babies: Cross-modal perception, imitation, and the mind of the preverbal infant. In N. Eilan, R. McCarthy, & B. Brewer (Eds.), *Spatial representation: Problems in philosophy and psychology* (pp. 219–235). Basil Blackwell.

Nanay, B. (2020). Molyneux's question and interpersonal variations in multimodal mental imagery among blind subjects. In B. Glenney (Ed.), *Molyneux's question* (pp. 257–263). Routledge. https://doi.org/10.4324/9780429020377

Nanay, B. (2023). Object files. In *Mental imagery: Philosophy, psychology, neuroscience* (online ed.). Oxford University Press. https://academic.oup.com/book/51673

Navarro-Guerrero, N., Toprak, S., Josifovski, J., et al. (2023). Visuo-haptic object perception for robots: An overview. *Autonomous Robots, 47*, 377–403. https://doi.org/10.48550/arXiv.2203.11544

Noë, A. (2004). *Action in perception*. MIT Press.

Norman, J. F., Crabtree, C. E., & Norman, H. F. (2006). Aging and the visual, haptic, and cross-modal perception of natural object shape. *Perception, 35*(10), 1383–1395. https://doi.org/10.1038/s41598-022-22141-z

O'Callaghan, C. (2014). Intermodal binding awareness. In D. J. Bennett & C. Hill (Eds.), *Sensory integration and the unity of consciousness*. Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/9780262027786.003.0004

O'Callaghan, C. (2015). The multisensory character of perception. *Journal of Philosophy, 112*(10), 551–569. https://doi.org/10.5840/jphil20151121035

O'Shaughnessy, B. (1989). The sense of touch. *Australasian Journal of Philosophy, 67*(1), 37–58. https://doi.org/10.1080/00048408912343671

Pascual-Leone, A., & Hamilton, R. (2001). The metamodal organization of the brain. *Progress in Brain Research, 134*, 427–445. https://doi.org/10.1016/S0079-6123(01)34028-1

Potter, M. C., Wyble, B., Hagmann, C. E., & McCourt, E. S. (2013). Detecting meaning in RSVP at 13 ms per picture. *Attention, Perception, & Psychophysics, 76*(2), 270–279. https://doi.org/10.3758/s13414-013-0605-z

Prather, S. C., Votaw, J. R., & Sathian, K. (2004). Task-specific recruitment of dorsal and ventral visual areas during tactile perception. *Neuropsychologia, 42*, 1079–1087. https://doi.org/10.1016/j.neuropsychologia.2003.12.013

Prinz, J. (2002). *Furnishing the mind: Concepts and their perceptual basis*. Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/3169.001.0001

Purpura, G., Cioni, G., & Tinelli, F. (2018). Development of visual-haptic transfer for object recognition in typical preschool and school-aged children. *Child Neuropsychology, 24*(5), 657–674. https://doi.org/10.1080/09297049.2017.1316974

Pylyshyn, Z. W. (2003). *Seeing and visualizing: It's not what you think*. Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/6137.001.0001

Pylyshyn, Z. (2009). The empirical case for bare demonstratives in vision. In R. J. Stainton & C. Viger (Eds.), *Compositionality, context and semantic values* (Studies in Linguistics and Philosophy, Vol. 85, pp. 215–239). Dordrecht: Springer. https://doi.org/10.1007/978-1-4020-8310-5_11

Quilty-Dunn, J. (2016). Iconicity and the format of perception. *Journal of Consciousness Studies, 23*(3–4), 255–263.

Quilty-Dunn, J. (2020). Perceptual pluralism. *Noûs, 54*(4), 807–838. https://doi.org/10.1111/nous.12285

Quilty-Dunn, J., & Green, E. J. (2023). Perceptual attribution and perceptual reference. *Philosophy and Phenomenological Research, 106*(2), 273–298. https://doi.org/10.1111/phpr.12847

Quilty-Dunn, J., Porot, N., & Mandelbaum, E. (2023). The best game in town: The reemergence of the language-of-thought hypothesis across the cognitive sciences. *Behavioral and Brain Sciences, 46*, e261. https://doi.org/10.1017/S0140525X22002849

Reid, J. M., Jacklin, D. L., & Winters, B. D. (2014). Delineating prefrontal cortex region contributions to crossmodal object recognition in rats. *Cerebral Cortex, 24*(8), 2108–2119. https://doi.org/10.1093/cercor/bht061

Reid, T. (1764/2007). *An inquiry into the human mind on the principles of common sense* (J. Bennett, Ed.). Edinburgh University Press. https://doi.org/10.1515/9781474471923

Rescorla, M. (2020). Perceptual co-reference. *Review of Philosophy and Psychology, 11*(3), 569–589. https://doi.org/10.1007/s13164-018-0411-6

Sathian, K., & Lacey, S. (2022). Cross-modal interactions of the tactile system. *Current Directions in Psychological Science, 31*(5), 411–418. https://doi.org/10.1177/09637214221101877

Scholl, B. J., & Pylyshyn, Z. W. (1999). Tracking multiple items through occlusion: Clues to visual objecthood. *Cognitive Psychology, 38*(2), 259–290. https://doi.org/10.1006/cogp.1998.0698

Schumacher, S., Burt de Perera, T., Thenert, J., & von der Emde, G. (2016). Cross-modal object recognition and dynamic weighting of sensory inputs in a fish. *Proceedings of the National Academy of Sciences of the United States of America, 113*(27), 7638–7643. https://doi.org/10.1073/pnas.1603120113

Schwenkler, J. (2012). On the matching of seen and felt shape by newly sighted subjects. *i-Perception, 3*(3), 186–188. https://doi.org/10.1068/i0525ic

Schwenkler, J. (2013). Do things look the way they feel? *Analysis, 73*(1), 86–96. https://doi.org/10.1093/analys/ans137

Serrahima, C. (2023). Sensory fields: The visual and the bodily. *Philosophical Studies, 180*(2), 679–700. https://doi.org/10.1007/s11098-022-01838-x

Smith, A. (1795). *Essays on philosophical subjects*. Dublin: Printed for Messrs. Wogan, Byrne, J. Moore, Colbert, Rice, W. Jones, Porter, and Folingsby.

Snow, J. C., Strother, L., & Humphreys, G. W. (2014). Haptic shape processing in visual cortex. *Journal of Cognitive Neuroscience, 26*(5), 1154–1167. https://doi.org/10.1162/jocn_a_00548

Solvi, C., Gutiérrez Al-Khudhairy, S., & Chittka, L. (2020). Bumble bees display cross-modal object recognition between visual and tactile senses. *Science, 367*(6480), 910–912. https://doi.org/10.1126/science.aay8064

Spence, C., & Di Stefano, N. (2024). What, if anything, can be considered an amodal sensory dimension? *Psychonomic Bulletin & Review, 31*(5), 1915–1933. https://doi.org/10.3758/s13423-023-02447-3

Spence, C., & Di Stefano, N. (2024). Old and new versions of the Molyneux question: A review of experimental answers. *Philosophy and the Mind Sciences, 5*, Article 11337. https://doi.org/10.33735/phimisci.2024.11337

Streri, A. (2012). Crossmodal interactions in the human newborn: New answers to Molyneux's question. In A. J. Bremner, D. J. Lewkowicz, & C. Spence (Eds.), *Multisensory development* (pp. 88–112). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199586059.003.0004

Treisman, A., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology, 14*(1), 107–141. https://doi.org/10.1016/0010-0285(82)90006-8

Versace, E., Freeland, L., & Emmerson, M. G. (2024). First-sight recognition of touched objects shows that chicks can solve Molyneux's problem. *Biology Letters, 20*, Article 20240025. https://doi.org/10.1098/rsbl.2024.0025

Winters, B. D., & Reid, J. M. (2010). A distributed cortical representation underlies crossmodal object recognition in rats. *Journal of Neuroscience, 30*(18), 6253–6261. https://doi.org/10.1523/JNEUROSCI.6073-09.2010

Zhang, M. M., Kennedy, M. D., Hsieh, M. A., & Daniilidis, K. (2016). A triangle histogram for object classification by tactile sensing. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (pp. 4931–4938). https://doi.org/10.1109/IROS.2016.7759724

Zhou, K., Luo, H., Zhou, T., Zhuo, Y., & Chen, L. (2010). Topological change disturbs object continuity in attentive tracking. *Proceedings of the National Academy of Sciences, 107*(49), 21920–21924. https://doi.org/10.1073/pnas.1010919108

SERGIO CERMEÑO-AÍNSA holds a PhD in Cognitive Science and Language from the Autonomous University of Barcelona and is a lecturer of Philosophy at the University of Girona (Spain). He has worked mainly on topics belonging to philosophy of mind, philosophy of perception and foundations of cognitive science.

ADDRESS: Department of Philosophy, Facultat de Lletres, University of Girona, Edifici Sant Domènec I, Pl. Ferrater i Mora, 1, Campus Barri Vell, Girona 17004, Spain. E-mail: elsergei@hotmail.com - ORCID: https://orcid.org/0000-0002-9598-8386

THEORIA
An International Journal for Theory, History and Foundations of Science