

Do things look the way they feel?

JOHN SCHWENKLER

1. A famous puzzle

Human perceivers can use vision and touch co-operatively in discerning the spatial features of nearby objects: for example, cross-modal comparisons by mature perceivers of the shapes of seen objects with felt ones are just as accurate as intramodal visual-visual and haptic-haptic identifications (Norman et al. 2004). In this respect we have a *multimodal* grasp of perceivable space: visual and tactile percepts can be compared in this way only because we recognize them as two modes of awareness of a single set of spatial features. But what accounts for this? Is it something we must learn to do, or is it due to something intrinsically common at the level of perceptual appearance?

To sharpen this question, compare the way we perceive events as ordered and extended in time. Anyone who can say whether two tones are played simultaneously or one before another, or which of the two is longer, should be able to make the same comparisons between a tone and a flash or a pulse on his skin. Intuitively, this is because there is a respect in which temporal features like succession and duration *appear* the same whether they are perceived through vision, touch or hearing: if you can identify these features by their appearance in one sensory modality, you are thereby equipped to recognize them in another. By contrast, though we can identify (e.g.) bananas by vision as well as gustation, this seems to be due just to associative learning: there is nothing in the look of a banana that tells you how it tastes, or vice versa. Which kind of explanation should we give of the ability to recognize spatial features through sight as well as touch?¹

An answer to this question would help in deciding between various philosophical theories of sense-perception. For example, consider Kant's claim that a single representation of space is 'the form of all appearances of outer sense' (1933: A26/B42). If something like this is right, then there should be a respect in which any given spatial feature should have the same appearance no matter which sensory modality it is perceived through: as J.J. Gibson suggests, in this respect these modalities should 'yield the *same* phenomenal experiences' (1962: 490), and so anyone who can identify spatial features through one sense will be prepared to do the same when they are perceived in a different modality.² By contrast, if there are subjects who can perceive spatially through

¹ Though my focus here will be on sight and touch, in principle this question could also be asked about hearing. For example, visual and auditory cues are integrated in the representation of stimulus location (Knudsen and Brainard 1995; Alais and Burr 2004) and perceived self-motion (Riecke, Völjamäe and Schulte-Pelkum 2009), information from touch and hearing can be integrated in the perception of surface texture (Guest et al. 2002), and human perceivers can rely on auditory cues in making judgments of object shape (Lakatos, McAdams and Caussé 1997; Kunkler-Peck and Turvey 2000) and size (Grassi 2005). Additionally, James et al. (2011) found that activation in the lateral occipitotemporal cortex (area LO), which has long been believed to be involved in visual and haptic shape processing, is heightened during auditory shape recognition tasks as well.

² This is also predicted by other philosophical theories of spatial perception that reject Kant's transcendental psychology. For example, Evans (1985: 389-90) argues that 'It is a consequence of the fact that the spatial content of auditory and tactual-kinaesthetic perceptions must be specified in the same, egocentric, terms, that perceptions from both systems will be used to build up a unitary picture of the world, and hence that spatial concepts applicable on the basis of one mode of perception must generalize to the other'. For a similar suggestion, see Noë (2004: 102, 110), who appeals to the existence of 'sensorimotor isomorphisms' between spatial experience in different modalities. I discuss Noë's view further in note 17 below.

sight and touch but cannot identify spatial features equally well in both modalities, then something in this Kantian position is wrong.³ In principle, we can make headway in evaluating such a doctrine by determining experimentally whether this prediction holds up.⁴

To see how we might do this, consider the question that William Molyneux posed to John Locke in 1688, and which Locke later included in the second edition of his *Essay Concerning Human Understanding* (II, ix, 8). Molyneux asked Locke whether a man born blind who had learned to recognize certain shapes on the basis of touch would, upon having his sight restored and the same shapes placed immediately before his eyes, be able to identify the shapes of the things he saw. Molyneux, and Locke following him, answered ‘Not’, arguing that such a person would have ‘not yet attained the Experience that what affects his touch so or so, must affect his sight so or so’ (Locke 1975: 146) – that is, they assumed that the visual and haptic appearances of solid shapes were intrinsically unlike one another, and could come to be associated only through perceptual learning.⁵ By contrast, if such a naïve subject could identify spatial features immediately on having them presented to his eyes, this would be consistent with the Kantian hypothesis outlined above.⁶

But we need to be cautious here. For obvious reasons, Molyneux’s statement of his puzzle assumes that the subject will be able to *see* well enough at the time of the crucial experiment to make out the shapes of the objects he is presented with, as we will not have learned much that is interesting if someone fails in the task due simply to visual deficiencies.⁷ And it is easy to see this requirement might end up making the puzzle into nothing more than a provocative thought-experiment.⁸ For example, if it is impossible for the human visual cortex to develop properly after the ‘critical period’ of early infancy or to become operational again after lying dormant for years, then Molyneux’s experiment cannot be run because the sufficient restoration of sight is simply impossible.⁹ Similarly, if the only way a newly sighted person can come to see adequately is through active, multimodal exploration of the world around him, then sight can be restored only in a way that would make the experiment uninformative.¹⁰ All this means that whether Molyneux’s puzzle can be resolved experimentally is itself an empirical matter: the possibility of an informative experiment of the sort he envisioned hinges on contingent

³ It is a Kantian position, but I will not insist that it was Kant’s. (Thanks to Angela Schwenkler for ensuring this.) For a more thorough study of how Kant thought about these issues, see Sassen 2004.

⁴ For another example of this strategy, see my (2012a). Importantly, as Robinson (1994: 208) notes, things do not go the other way: if naïve subjects *can* make knowledgeable intermodal comparisons immediately, this may be due to ‘hard-wiring’ rather than intrinsically shared appearances.

⁵ Once again, I have no pretension to historical accuracy, as Locke’s actual position was probably more complex than this; here see Bruno and Mandelbaum 2010.

⁶ Another way psychologists have tried to answer this question is by looking for signs of cross-modal recognition in human infants (e.g. Meltzoff and Borton 1979). However, it is possible that what drives these effects is something other than shared appearances at the level of perceptual experience; it could also be due to connections at the sub-personal level. In this respect it seems preferable to test mature subjects’ capacities for explicit recognition of spatial features.

⁷ A point stressed e.g. by Evans (1985: 380).

⁸ As Degenaar (1996: 25) notes, this was how the first generation of philosophers to respond to Molyneux’s puzzle understood it.

⁹ The view of Degenaar (1996: 132) and Gallagher (2005: 165-68), for example.

¹⁰ The view of Noë (2004: 102-3).

facts about the human visual system.¹¹ But recent research on the visual capacities of the newly sighted gives reason for optimism on this score, as even congenitally blind individuals have been made to see in a way that supplies considerable visual capacity immediately, and a wider battery of visual skills within a matter of days or weeks.¹² However, I will argue below that the results of this research have failed thus far to resolve Molyneux's puzzle unambiguously, before proposing a way to improve it.

2. Recent research

In order to accomplish what it aims, an experiment run along Molyneux's lines must meet two main conditions: first, its subjects must not have formed any learned associations between spatial appearances in the relevant sense-modalities; and second, these subjects must not be too deficient in the senses being employed. Together, these conditions capture what Richard Held identifies as 'the stringent requirements for definitively answering [Molyneux's] question':

First of all, the [patient's] blindness must have been verifiably congenital and continuous. Otherwise there may have been an opportunity for acquisition of the crossmodal transfer through visual experience whose exclusion is the purpose of testing the previously blind. ... After successful surgery the patient must exhibit acuity sufficient to discriminate visually among the objects used for testing. ... Post-op testing should begin as soon after surgery as possible – ideally when bandages are first removed. Crossmodal recognition has been reported in some cases long after surgery. But if it has not been tested immediately we cannot exclude the possibility of acquisition by experience ... (2009: 585)

The bulk of this quotation stresses the importance of ensuring that the subject has no opportunity, either before or after the surgery, to form associative connections between sight and touch. And in the third sentence, Held suggests a way of testing the adequacy of the subject's sight: if the patient can distinguish the stimulus objects visually, then – it is assumed – any failure to match seen and felt shapes cannot be the product of a purely visual deficit. (I will return shortly to issues that complicate this assumption.) Having ensured these things, we are supposed to be able to run Molyneux's experiment, and determine how sight and touch come to be connected.

With this in mind, in a recent study Held and colleagues presented five newly sighted individuals with 20 pairs of stimulus objects, each constructed from Lego blocks (see Figure 1) and 'large enough ... to sidestep any acuity limitations of the subjects' (Held et al. 2011: 551). Presentation of the stimuli involved the display of a single shape, which was then joined by two other shapes, one identical to the first and the other different from it, with the instruction to identify which of the latter two shapes matched the original one. These stimuli were presented in three different manners: either all the shapes were presented to touch alone (the 'touch-to-touch' task, *TT*), all were presented to vision alone ('vision-to-vision', *VV*), or the original shape was presented to touch and then the subsequent pair to vision ('touch-to-vision', *TV*). The crucial question was whether recognition of the original shape was worse in the touch-to-vision task than in the other two.

¹¹ However, the impossibility of answering the question experimentally in the way Molyneux envisioned it would not entail that it can be given no empirical resolution at all; here see Jacomuzzi, Kobau and Bruno 2003.

¹² See Fine et al. 2003, Maurer et al. 2005, Ostrovsky et al. 2006, Mandavilli 2006 and Thomas 2011.

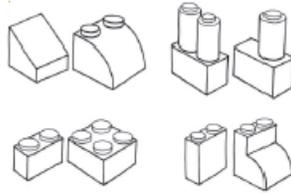


Figure 1: Sample stimulus pairs from the Held et al. (2011) study. Subjects were shown a single object from any given pair, which was then joined by two other objects, one the same as the first and the other different from it. They had to identify which of the latter two objects matched the first one.

Strikingly, despite very high performance in the two unimodal tasks (mean 98% in TT, and mean 92% in VV), performance in the cross-modal task was barely above chance (mean 58%). That is, despite being able to match seen shapes with seen ones and felt with felt, newly-sighted individuals were entirely inaccurate in their cross-modal comparisons between vision and touch. Within five days after the original test, however, and ‘given only natural real-world treatment’ without any explicit training (Held et al. 2011: 552), subjects’ performance in the TV task using novel but similar stimuli averaged near 80%. Thus, the authors conclude, ‘the answer to Molyneux’s question is likely negative’:

The newly sighted subjects did not exhibit an immediate transfer of their tactile shape knowledge to the visual domain. This finding has important implications for bimodal perception. Whatever linkage between vision and touch may pre-exist concomitant exposure of both senses, it is insufficient for reconciling the identity of the separate sensory representations. (ibid.)

In short, Held and colleagues contend that the perception of a given shape through sight and touch does not guarantee the ability to recognize these perceptual appearances as appearances of the same shape. This contradicts the idea that the visual and tactile appearances of spatial features are intrinsically the same.

3. *A sceptical assessment*

Immediately upon being made to see, Held and colleagues’ subjects could not match the shapes they saw with ones they also perceived through touch. But does the fact that the subjects were able to match seen shapes with seen ones in the vision-to-vision task guarantee that their failure in the touch-to-vision task cannot have been rooted in a purely visual deficit?

Further reflection on the nature of the tasks should lead us to doubt this assumption. In the VV task, subjects needed only to make gross discriminations based on the overall appearance of the stimuli, which were presented from a single viewing angle. Intuitively, this can be done by attending to low-level visual features like colour, shadow and approximate overall contours: think for example of what it is like to distinguish objects seen at a far distance, without being able really to make them out. By contrast, such a crude strategy would not suffice for the cross-modal task, which made low-level visual cues irrelevant and demanded robust shape representations that could be compared across modalities.¹³ Given this difference, the subjects’ evident ability to discriminate objects

¹³ I make no claim as to the precise nature of these representations, e.g. whether they are structural or image-based, viewpoint-dependent or viewpoint-invariant; for a review of the possible options with respect to visual object recognition, see Barenholz and Tarr 2007. On the

visually does not guarantee that their capacity for visual form perception sufficed for an experimental resolution of Molyneux's puzzle.

This sceptical analysis is supported by other experimental work carried out by the same group. Between two weeks and three months after surgical restoration of their vision, three patients were shown images of 'three dimensional shapes, such as cubes and pyramids, ... with surfaces of different luminance consistent with lighting and shadows' (Ostrovsky et al. 2009: 1486). In this condition, 'the recently treated subjects reported perceiving multiple objects, one corresponding to each facet. They were unable to integrate the facets into the percept of a single three-dimensional objects' (ibid.). Nor did they fare any better with photographic images of common objects: they could not name the stimuli, and when instructed 'to point to objects in the images and also to indicate their extent', the subjects 'pointed to regions of different hues and luminances as distinct objects. This approach greatly oversegmented the images and partitioned them into meaningless regions, which would be unstable across different views and uninformative regarding object identity' (ibid.: 1487). At first glance, this might seem to leave quite inexplicable the success of Held and colleagues' subjects in the VV task: yet as we have seen, success in that task did not really demand 'percepts of single three-dimensional objects' at all, but only attention to low-level features.¹⁴ But the TV task required more than this, and so the subjects' failure in may have been due to the inadequacy of their newly restored sight.

Put differently, the problem is that while success in the VV task is clearly *necessary* to ensure the adequacy of the subjects' sight, it alone does not suffice for this.¹⁵ In addition, we would need evidence that Held and colleagues' subjects could not only discriminate between objects displayed from a single viewing angle, but also re-identify those objects when their visual perspective on them was changed: for only so would it be shown that they were visually sensitive to the perspective-invariant spatial features that would underwrite a capacity for cross-modal matching, as opposed to other cues that sufficed only for intramodal discrimination. However, the evidence cited above makes it appear overwhelmingly unlikely that they were able to do this, at least given the relative complexity of the shapes and the limited visual information that was available to them. (I will return to this last point below.) Therefore we should conclude that Held and colleagues' subjects failed to match seen shapes with felt ones simply because they could not form the requisite visual representations of those shapes in the first place.

4. Further directions

Despite their efforts, Held and colleagues' attempt to resolve Molyneux's puzzle was unsuccessful: they did not ensure that their subjects could form robust representations of visual shape, and other experimental work on the restoration of sight suggests that the subjects cannot have done this within the conditions of the experiment. However, it would be hasty to conclude from this that the visual capacities of the newly sighted are

apparent viewpoint-independency of cross-modal object recognition in normal subjects, see Lacey, Peters and Sathian 2007, Lacey et al. 2009, and Lacey, Hall and Sathian 2010.

¹⁴ Held and colleagues do cite this study, but only to challenge an account of their subjects' rapid improvement in the TV task as the result of 'a rapid increase in the visual ability to create a three-dimensional representation, thus allowing for a more accurate mapping between haptic structures and visual ones' (2011: 552). This may be right, but it overlooks the more fundamental problem that these results raise for the interpretation of their data.

¹⁵ Compare Evans' remark that 'a capacity to make gross "same/different" judgments is far from establishing visual perception of [a] figure' (1985: 380).

necessarily too limited for Molyneux's puzzle to be subject to straightforward experimental resolution.

To see why not, note that in contrast to the tactile displays, where subjects could manipulate the objects they held in their hands, the visual stimuli presented by Held and colleagues were always motionless, and subjects were permitted only 'to adjust their distance or viewpoint while remaining seated in front of the presentation table' (Held et al. 2011: suppl. info.). In everyday life, the visual perception of complex three-dimensional shapes is often quite unlike this: just as you perceive a shape through touch by running your hands all around it,¹⁶ so visual shape perception frequently involves more than just looking at an object from a single angle. Despite the 'simultaneity' of the spatial information available to vision, watching an object as it moves with respect to one – or as one moves with respect to it – can bring into view features of it that would otherwise have been hidden, and makes available more information concerning how its surfaces are oriented in depth. (Indeed, in this respect visual shape perception is actually *less* simultaneous than is the perception of shape through touch: you may see something from only one side at a time, but touch an object from several angles at once.) Yet Held and colleagues barred their subjects from drawing on this sort of information, and limited them to whatever could be seen of the visual stimuli from just a single side. This kept them from accessing the richer array of spatial cues that a changing perspective on the object would have supplied, while at the same time making it especially easy to use low-level features in the vision-to-vision task, since they could be sure that they would encounter each object from the same viewing angle.

Here one might object that newly sighted individuals probably lack the perceptual skills necessary to draw on such dynamic visual information,¹⁷ but in fact this may not be the case. For example, five months post-surgery patient MM was generally insensitive to visual perspective cues and would identify the image of a stationary Necker cube as 'a square with lines' (Fine et al. 2003: 915); however, when motion-in-depth was simulated he 'immediately saw a cube' (ibid.). Similarly, Ostrovsky et al. found that their subjects were more able to recognize photographic images of objects that in ordinary life are more likely to be seen in motion, suggesting that 'motion of objects helps bind their constituent regions into cohesive representations, which can then be used to recognize instances in new inputs that may be static' (2009: 1489). It would, then, be better to re-run Held and colleagues' experiment with the stimulus objects made to move, and/or the subjects moved or permitted to move with respect to them. This would increase the visual information available to the perceivers, and could improve their ability to bind together the facets of the objects and make out their 3D shapes.¹⁸ As it is, the study fails to resolve the questions it aims to address, and only shows how badly newly sighted individuals do at visually discerning the three-dimensional shape of a motionless object that they are permitted to see from just one angle – and this much should be unsurprising, especially

¹⁶ What Gibson (1962) calls 'active touch'. On the multidimensionality of haptic object perception, see Klatzky and Lederman 2011.

¹⁷ This seems to be the view of Noë, who claims that newly sighted individuals suffer from 'experiential blindness', or blindness 'due not to the absence of sensation or sensitivity, but rather to the person's (or animal's) inability to integrate sensory stimulation with patterns of movement and thought' (2004: 4) – an ability that he supposes can be acquired only through a process of sensorimotor exploration. But the evidence reviewed here, especially those papers cited in note 12 above, is in tension with this supposition. If Noë's overall theory is correct, then some sensorimotor know-how may be innate.

¹⁸ For some other possible ways to improve the study, see the final section of my (2012b), whose argument this paper extends.

when our attention has been called to the probing exploration that visual perception frequently involves.¹⁹

Mount St. Mary's University
Emmitsburg, MD 21727, USA
schwenkler@msmary.edu

References

- Alais, D. and D. Burr. 2004. The ventriloquist effect results from near-optimal bimodal integration. *Current Biology* 14: 257-62.
- Barenholtz, E. and M. J. Tarr. 2007. Reconsidering the role of structure in vision. In *Categories in Use*, ed. A. Markman and B. Ross 157-80. San Diego: Academic Press.
- Bruno, M. and E. Mandelbaum. 2010. Locke's answer to Molyneux's thought experiment. *History of Philosophy Quarterly* 27: 165-80.
- Degenaar, M. 1996. *Molyneux's Problem*. Trans. M.J. Collins. Dordrecht: Kluwer.
- Evans, G. 1985. Molyneux's question. In his *Collected Papers*, 364-99. Oxford: Clarendon Press.
- Fine, I., Wade, A. R., Brewer, A. A., May, M. G., Goodman, D. F., Boynton, G. M., Wandell, B. A. and D. I. A. MacLeod. 2003. Long-term deprivation affects visual perception and cortex. *Nature Neuroscience* 6: 915-16.
- Gallagher, S. 2005. *How the Body Shapes the Mind*. New York: Oxford University Press.
- Gibson, J. J. 1962. Observations on active touch. *Psychological Review* 69: 477-91.
- Grassi, M. 2005. Do we hear size or sound? Balls dropped on plates. *Attention, Perception, & Psychophysics* 67: 274-84.
- Guest, S., Catmur, C., Lloyd, D. and C. Spence. 2002. Audiotactile interactions in roughness perception. *Experimental Brain Research* 146: 161-71.
- Held, R. 2009. Visual-haptic mapping and the origin of crossmodal identity. *Optometry & Vision Science* 86: 595-98.
- Held, R., Ostrovsky, Y., de Gelder, B., Gandhi, T., Ganesh S., Mathur, U. and P. Sinha. 2011. The newly sighted fail to match seen shape with felt. *Nature Neuroscience* 14: 551-53.
- Jacomuzzi, A., Kobau, P. and N. Bruno. 2003. Molyneux's question redux. *Phenomenology and the Cognitive Sciences* 2: 255-80.
- James, T. W., Stevenson, R. A., Kim, S., VanDerKlok, R. M. and K. H. James. 2011. Shape from sound: evidence for a shape operator in the lateral occipital cortex. *Neuropsychologia* 49: 1807-15.
- Kant, I. 1933. *Critique of Pure Reason*. Trans. N. Kemp Smith. New York: Palgrave.
- Klatzky, R. L. and S. J. Lederman. 2011. Haptic object perception: spatial dimensionality and relation to vision. *Philosophical Transactions of the Royal Society B* 366: 3097-105.
- Knudsen, E. I. and M. S. Brainard. 1995. Creating a unified representation of visual and auditory space in the brain. *Annual Review of Neuroscience* 18: 19-43.
- Kunkler-Peck, A. J. and M. T. Turvey. 2000. Hearing shape. *Journal of Experimental Psychology: Human Perception & Performance* 26: 279-94.

¹⁹ I presented versions of this material at Mount St. Mary's University, at the 2012 meeting of the Southern Society for Philosophy and Psychology, and in a talk I gave to Dr. Bill Eaton's Irish Philosophy class in July 2012. Thanks to audiences on those occasions, and especially to Jacob Berger, Robert Briscoe, Mike Bruno, Pat Churchland, Kevin Connolly, Pete Mandik, Mohan Matthen, Greg Murry, Angela Schwenkler, James Stazicker and Arnold Trehub for comments and discussion.

- Lacey, S., Peters, A. and K. Sathian. 2007. Cross-modal object recognition is viewpoint-independent. *PLoS ONE* 2: e890.
- Lacey, S., Pappas, M., Kreps, A., Lee, K. and K. Sathian. 2009. Perceptual learning of view-independence in visuo-haptic object representations. *Experimental Brain Research* 198: 329-37.
- Lacey, S., Hall, J. and K. Sathian. 2010. Are surface properties integrated into visuohaptic object representations? *European Journal of Neuroscience* 31: 1882-88.
- Lakatos, S., McAdams, S. and R. Caussé. 1997. The representation of auditory source characteristics: simple geometric form. *Perception & Psychophysics* 59: 1180-90.
- Locke, J. 1975. *An Essay Concerning Human Understanding*. Ed. P.H. Nidditch. New York: Oxford University Press.
- Mandavilli, A. 2006. Look and learn. *Nature* 441: 271-72.
- Maurer, D., Lewis, T. L. and C. J. Mondloch. 2005. Missing sights: consequences for visual development. *Trends in Cognitive Sciences* 9: 144-51.
- Meltzoff, A. N. and R. W. Borton. 1979. Intermodal matching by human neonates. *Nature* 282: 403-4.
- Morgan, M. 1977. *Molyneux's Question*. Dordrecht: Kluwer.
- Noë, A. 2004. *Action in Perception*. Cambridge: The MIT Press.
- Norman, J. F., Norman, H. F., Clayton, A. M., Lianekhammy, J. and Zielke, G. 2004. The visual and haptic perception of natural object shape. *Perception and Psychophysics* 66: 342-51.
- Ostrovsky, Y., Andalman, A. and P. Sinha. 2006. Vision following extended cortical blindness. *Psychological Science* 17: 1009-14.
- Ostrovsky, Y., Meyers, E., Ganesh, S., Mathur, U. and P. Sinha. 2009. Visual parsing after recovery from blindness. *Psychological Science* 20: 1484-91.
- Riecke, B. E., Våljamäe, A. and J. Schulte-Pelkum. 2009. Moving sounds enhance the visually-induced self-motion illusion (circular vection) in virtual reality. *ACM Transactions on Applied Perception* 6: article no. 7.
- Robinson, H. 1994. *Perception*. New York: Routledge.
- Sassen, B. 2004. Kant on Molyneux's problem. *British Journal for the History of Philosophy* 12: 471-85.
- Schwenkler, J. 2012a. Does visual spatial awareness require the visual awareness of space? *Mind and Language* 27: 308-29.
- . 2012b. On the matching of seen and felt shapes by newly sighted subjects. *i-Perception* 3: 186-88.
- Thomas, S. 2011. Project Prakash: challenging the critical period. *Yale Journal of Biology and Medicine* 84: 483-85.