# ACTIVE PERCEPTION AND THE REPRESENTATION OF SPACE

## Mohan Matthen

In the *Transcendental Aesthetic* of the *Critique of Pure Reason*, Kant famously asserts that (a) space is *the* form of outer, and time *the* form of inner, intuition. He conceives of space, in particular, as specific to no particular sense modality, but rather as common to any receptive modality *qua* receptive: it is (b) “the condition of the possibility of appearance,” (B39) (c) “prior to any perception of an object” (B41), (d) “the subjective condition of sensibility, under which alone outer intuition is possible for us” (B42).

Rephrasing the above in somewhat more contemporary terms, Kant propounds the following *Pre-Modality Thesis* (as I shall call it):

1. [*Commonality*] There is a single representation of space and of time common to all modalities. (a)
2. [*Aprioricity*]These representations are templates for the perceived spatiotemporal ordering of objects and are not derived from information that impinges on the senses. (b, c)
3. [*No Privileged Modality*] No single modality has priority with regard to these templates. Thus, the common representation of space is not simply projected from vision or touch to the other modalities. (c)
4. [*Outerness*]“Appearances” purport to be about the world outside the mind by virtue of being arrayed in the metric denoted by the pre-modal representation of space. (d)

Kant’s position about space was surely controversial at a time when Molyneux’s question was thought to be deeply revealing about the epistemic significance of experiences provided by the sensory modalities (Sassen 2004). Sense impressions of different modalities differ in (what we would now call) phenomenal character. And at the time it was assumed that if any two sense impressions were phenomenally different then they must be different with respect to their cognitive significance. Thus it was thought to follow that even if sense impressions of different modalities purport to be of the same quality—for example, the same shape—they were nevertheless different with respect to their cognitive significance. This was the tide that Kant swam against. His approach continues to be controversial today.

The theses articulated above are, of course, only a part of Kant’s theory of space. I shall not be concerned here with another closely associated position that he articulates in the *Transcendental Aesthetic*: the idea that because it is known *a priori*, space is a *mental* template, which does not characterize things in themselves. My general attitude is that this is a *non sequitur*. One could grant that the perceptual representations of space and time are *a priori* mental constructs, but still insist that space and time are objective entities represented by perception, whether veridically or not.(Evolution, or God, could have ensured a large degree of convergence between space and time and their representations.) In this paper, I am interested primarily in what Kant had to say about the representations of space and time. Accordingly, I shall take a realist stance about space and time themselves, but offer no more than a cursory defence of this position.

My main aim in this paper is to defend Kant’s Pre-Modality thesis as summarized above. However, I will not devote much attention to Outerness. Though it forms an integral part of Kant’s thinking about this subject, I don’t think that he formulates the Outerness thesis in exactly the right way—see note 2. Outerness appears to have a good deal of merit and deserves further investigation, but it goes well beyond my aims here to construct a corrected version that would be worth defending.

### Kant’s Arguments

Kant’s Pre-Modality thesis has fallen out of sight in the philosophical discussion of space, but it is important and original and, in my view, essentially correct. In the *Transcendental* Aesthetic, we find hints of two arguments in support of it. In this opening section, I shall present these arguments and argue that they are inadequate. My aim in the paper is to offer a different argument in support of Pre-Modality—a Common Measure argument, as I shall call it.

1. Outerness and Pre-Modality

Kant seems to have thought, first, that the Outerness thesis implies the rest of the Pre-Modality thesis. Before I attempt to reconstruct his argument, let me explain what he means by Outerness.

Kant claims that things appear to be outside the mind because the senses locate them in space. In other words, he holds (in something of an echo of Descartes’ conception of matter) that the perception of things as spatially ordered is constitutive of perceiving them as existing objectively—that is, independently of the subject’s mental domain.[[1]](#footnote-1)

It is instructive to compare Kant’s position here with that of Hume:

. . . tho’ every impression and idea we remember be consider’d as existent, the idea of existence is not deriv’d from any particular impression. (*Treatise* I.II.vi)

According to Hume, the perceptual image, or idea, of something is the same whether one thinks of it as existent or not. In other words, when I imagine a scene as non-existent, what I imagine is qualitatively the same as when I imagine the same scene as existent. Since supposing things to be existent does not change how they appear to the senses, existence must be an extra-sensory idea. Kant agrees with this: he famously holds that existence is not a ‘predicate’.

When I think a thing, through whichever and however many predicates I like, [even in its thoroughgoing determination], not the least bit gets added to the thing when I posit in addition that this thing is.” (*Critique of Pure Reason*, B628)

However, Kant introduces a new and different notion in the *Transcendental Aesthetic*: the perceptual appearance of being “outer” (i.e., outside the mind, or objective). Hume would presumably have held that the notion of outerness too is not “deriv’d from any impression”: he would have held that for any impression that the subject takes as possessing an external correlate, there could be an exactly similar impression that the subject takes as possessing *no* external correlate. Thus, he would have held, no sensory quality differentiates outer and inner intuition. Kant demurs. He thinks that when something looks to be located in space, it appears as if it exists outside the mind. For Kant, in other words, the perceptual appearance of spatial location *is* the appearance of outerness.[[2]](#footnote-2)

Does this notion of space as the mark of the outer yield an argument for Pre-Modality? Perhaps Kant thought so, for he writes:

 . . . one can represent only one space, and if one speaks of many spaces, one thereby understands only parts of one and the same unique space. These parts cannot precede the one all-embracing space as being, as it were, constituents out of which it can be composed, but can only be thought as *in it*. It is essentially one; the manifold in it, and therefore also the general concept of spaces, depends solely on limitations. *It follows from this that an a priori intuition (which is not empirical) underlies all concepts of space.* (B39; emphasis added)

This passage makes a bold and important claim that goes well beyond the Outerness thesis (which I dispute in note 2). The further claim he now makes is that since “one can represent only one space,” one cannot sense two objects as outer without sensing that they are located in the same space. He claims, in other words, that to sense something as external is *eo ipso* to sense it as existing in the same space as everything else sensed as external. Call this the one-sensory-space proposition (1SS). This, Kant maintains, is an *a priori* intuition. Clearly, this position is contestable. For one might think that the “one all-embracing space” is given to us by reason, and not embedded in perceptual experience *a priori*.

Let’s grant Kant 1SS for the sake of argument. The question arises: Does it imply that there is only one intuition of space common to all the modalities? Kant’s line of thought seems to run as follows.

1. Each of the modalities locates things in space, and thereby as existing objectively.
2. But “one can represent only one space.”
3. Therefore, there is a unique representation (or “intuition”) of space in which every modality locates things.[[3]](#footnote-3)
4. [Aprioricity and No Privileged Modality are reasonable assumptions for a modality-independent representation of space: see the following subsection.]

Again, this is not a good argument. For though we might represent only one space, it does not follow that there is just one representation of it. (This, exactly, is the point of the different phenomenal character argument given in the introductory section of this paper.) Grant that our representation of things as existing objectively implies that they exist side by side in one and the same space. It may nonetheless be that each modality represents space in its own way, and it may not be evident to us how these diverse representations of space match up (Hopkins 2005).

To elaborate, consider touch and vision.

I may see my finger, and feel a pin pressing down on it. The finger looks as if it is in front of the TV, but the pin does not feel as if it is in front of the TV. The TV, the pin, and the feel of the pin are all spatially located, one by vision, one by both vision and touch, and the last by touch. But touch does not appear to represent space in a way that makes the relationship of the three objects evident.

Place your left and right index fingers on the wall in front of you. Now, close your eyes and try to touch your nose to the wall so that it lies on a straight line between the two fingers. It is very hard to do this. (I did it in the shower, so that my digits and nose left wet imprints on a glass divider. Usually, the nose came out higher, though it was possible to get it right by attending carefully.) Again, it seems as if touch does not immediately represent how parts of your body are related to one another relative to the visual representation.

These failures, or limitations, of spatial matching suggest, at the very least, that the perceptual matching of visual and haptic representations of space is imperfect.[[4]](#footnote-4) On the other hand, it seems that the match between vision and audition is smooth: if you hear a sound coming from a particular place, you can spatially relate it to things you see. In short, there are cross-modal matches of spatial representation as well as limitations of cross-modal matching. And this shows that one cannot argue for any of this *a priori*, as Kant seems to do.[[5]](#footnote-5)

1. Cognitive Necessity and Pre-Modality

A second argument appeals to cognitive necessity. We cannot conceive of *perceived* space as other than three-dimensional and Euclidean, Kant says;[[6]](#footnote-6) these features of space are “apodeictic.” But nothing that we learn through the senses is apodeictic: every such proposition can always be conceived to be otherwise. It follows that the three-dimensional Euclidean character of space must be known *a priori*, not inferred from the data provided by sensation. (Actual measurements tend to show that haptic representations of space depart significantly from Euclidean structure: see, for example, Kappers and Koenderink 1999. So even if Kant is right about cognitive necessity, he is wrong the content of these cognitive necessities, and in particular about perceived space being Euclidean.)

Kant rightly thought that this argument supports the idea that our perceptual representation of space is not arrived at discursively, i.e. by induction, but is rather an “*a priori* intuition.”(It’s an “intuition” because space is apparently an object of the senses.) Of course, this argument does not show that there is no *a posteriori* component in spatial perceptions. Obviously, I can only know by looking (or by feeling by touch) that my pen is to the right of my keyboard, not to the left. Nevertheless, there are apodeictic aspects of my perceptions of spatial arrangement. For instance, I cannot perceive something as being external without perceiving it as being somewhere. Moreover, certain spatial relations are apodeictic despite being perceived: for example, that the straight path from my pen to my keyboard is shorter than the more circuitous route via the computer monitor. It is implicit in my perception of these three things that anything that looks like *that* has the appearance of outerness and, again, that *in virtue of appearing straight*, a line appears to be the shortest distance between two points.[[7]](#footnote-7) This kind of idea is constitutive of what Kant calls the *a priori* intuition of space.

Thus understood, however, our ideas about spatiotemporal structure are not our only *a priori* intuitions. Consider the representation of colour. The visual awareness of colour also carries apodeictic awareness of certain relations among the colours. For example, one cannot experience anything as orangewithout experiencing it as containing a mixture of *yellowish* and *reddish* components.In exactly the same way as the triangle inequality, it is necessary that all orange things be colour-mixed in this way.Similarly, one experiences certain shades as simple and others (such as *orange*)as composed of certain colour components. Again, it is often suggested (by Wittgenstein, for example, in the *Remarks on Colour*)that our experience of colour is, as such, sufficient to assure us that nothing could be reddish and greenish at the same time.Awareness of this and other aspects of the “similarity space” of colour is, as it were, “the subjective condition of the possibility of visual appearance of colour.”[[8]](#footnote-8)

Kant discovered an important phenomenon: that certain aspects of *individual* spatial perceptions are imbued by feelings of cognitive necessity, that is, by the “intuition” that things cannot occupy certain spatial relations. But he did not appreciate that this is true of the perception of any quality. In the case of colour perception, this feeling is the product of a certain kind of formatting. This does not, as Kant mistakenly thought, imply that the *qualitative fields* in which these cognitive necessities are found must be mental constructs. One could, consistently with the formatting supposition, insist that what we perceive as colour is real: for example, that it is surface reflectance.[[9]](#footnote-9) The colour vision system takes in particular instances of this real field and records them in a certain format: simply speaking, it *encodes and displays* every reflectance as a combination of three values on the orthogonal axes of blue-yellow, red-green, and dark-light. (This format is not dictated by the physical nature of colour.) The cognitive necessities of colour perception arise from the structure of this coding. For example, the impossibility of something being both reddish and greenish arises from the fact that these are opposite ends of a single coding axis. Similar things can be said of other perceptual qualities. Psychophysicists have constructed similarity spaces for pitch and flavour. Arguably, even primary qualities such as shape and motion present themselves in similarity relations that reflect more how they are perceptually coded than anything about the physical character of what is perceived.

In similar fashion, suppose that space is real. It may nevertheless be that our perceptual system encodes the spatial locations of things in a Euclidean framework. For example, it receives contingent locational data from my pen and my keyboard and represents these as points in a Euclidean frame. What we should take from Kant’s argument is not that space itself is perceptually constructed, but that the *a priori* intuition of space arises from a perceptual encoding scheme.[[10]](#footnote-10)

None of this helps with the Pre-Modality thesis, however. Why should Kant have thought that the perceptual representation of space is not specific to the sense modalities? Possibly, his thinking may have gone something like this:

1. (i) The perceptual representation is *cognitively necessary*, so it is *a priori*.
2. (ii) *A priori* representations do not depend on sensory input.
3. (iii) Representations that do not depend on sensory input are not modality-specific.
4. Therefore:
5. (iv) The perceptual representation of space is not modality-specific. (Commonality)
6. (v) [No Privileged Modality follows from (ii) and (iii); Aprioricity from (i) and (ii).]

This argument is defeated by the above diagnosis of cognitive necessity in perception. It lays cognitive necessity and aprioricity at the feet of perceptual coding. But, as the example of colour shows, perceptual coding can be modality-specific.

Kant’s argument points to certain structural “*a priori* intuitions” of space. Plausibly, these arise from the manner in which spatial qualities are encoded by perceptual systems. One can see this as a limited defence of the Aprioricity thesis stated above. But the inference from the Aprioricity of such intuitions to their Pre-Modality is impermissible.

### The Pre-Modality of Time Perception

At first sight, time is multimodal in the respect just discussed. Events sensed in one modality line up temporally both with one another and with events sensed in all other modalities. When one witnesses a musical song-and-dance act, the music one *hears* temporally lines up with the dancing one *sees*; when one sees an event and is jabbed on the finger with a pin, the events are generally sensed as standing in a particular temporal ordering. More generally:

To be aware of an event E1 in any modality and aware of event E2 in any modality (including non-external modalities) is to be aware of an apparent temporal relationship between E1 and E2 (provided that E1 and E2 are close enough in time that their temporal relationship can be sensed, rather than calculated or inferred.)[[11]](#footnote-11)

All awareness is *inter alia* awareness of time. So it seems that time is a common measure of events sensed across modalities.

The temporal matching of objects of experience seems bound up with what Ian Phillips (forthcoming) has called a “naïve view” (which he vigorously defends). In his words:

When all goes well, your stream of consciousness simply inherits the temporal structure of the events which are its contents. . . As a result the temporal structure of experience matches the temporal structure of its objects. In cases of illusion, it is as if this is so.

The idea is that when, for example, you experience a song as taking a certain amount of time, it is by your experience of the song taking that amount of time. Similarly, when you experience event *E1* as occurring earlier than event *E2*, it is by your experience of event *E1* occurring earlier than your experience of event *E2*. In short, the naïve view can be summed up in the following principle:

Exportation of Temporal Operators (First Pass)

A subject perceptually experiences events E1 and E2 as standing in temporal relation R if and only if her perceptual experiences of E1 and E2 stand in temporal relation R, subject to the proviso that she perceptually experiences E1, E2, [[12]](#footnote-12) and the temporal relation R.[[13]](#footnote-13)

For example, *S* experiences *E1* occurring *a little earlier* than *E2* if and only *S*’s experience of *E1* occurs a little earlier than her experience of *E2*.

On the face of it, the Exportation Principle seems to offer us a very simple explanation of the temporal synchronization of the song and dance that one hears and sees, one that altogether bypasses the *representation* of time. The dance and the accompanying music seem simultaneous because the experiences take place at the same time. The organizing matrix that gives events their perceivedtemporal order is *time itself*. Since time is real, not merely a perceptual representation, the representation of temporal order is extra-sensory. (Of course, Kant would disagree with this last statement.)

Actually, things are more complicated than this. Temporal sequencing in perception is the result of a complex computational process. As an illustration of why such a process is needed, and why it is quite complex, consider first what happens when somebody touches your toe and your nose simultaneously. Normally, the tactile experiences are felt to be simultaneous; that is, you have the capacity to judge whether the two touches were simultaneous or not. Yet, the neural signal from the toe, having a much longer distance to travel, arrives at the brain later than the signal from the nose. The brain must, for this reason, tag signals by their point of origin, holdthem in abeyance until signals from other parts of the body arrive (this buffer is around 80 ms long, as it turns out), and then sort out temporal relations within the buffer according to point of origin (Eagleman 2010).

Temporal illusions are further evidence of this. David Eagleman (2009) writes:

Imagine that every time you press a key, you cause a brief flash of light. Now imagine we sneakily inject a tiny delay (say, two hundred milliseconds) between your key-press and the subsequent flash. You may not even be aware of the small, extra delay. However, if we suddenly remove the delay, you will now believe that the flash occurred *before* your key-press.

Illusions are a much used method of getting the brain to reveal its computional methods: this is a case of association being used to compute temporal sequence.

Further, there is no *a priori* reason to believe that when a subject experiences two events as standing in temporal relation *R*, her experiences of these events actually stand in relation *R*. They may simply seem to do so. For it is natural to think that a subject’s experience of two events standing in temporal relation *R* is separate from her experiences of the two events; it is a meta-experience, an experience of experiences. And this is the most natural interpretation of what happens in the temporal illusions: in Eagleman’s example above, it is natural to suppose that though the experiencesof the flash (*F*) and of the key-press (*K*) are virtually simultaneous, there is a meta-experience of the experience of *F* being earlier than the experience of *K*. As William James (1890) wrote: “The mental stream, feeling itself, must feel the time-relations of its own states.” (628)

For these reasons, Phillips (2008) is surely wrong to say that:

. . . we cannot make sense of the idea that experience systematically seems to one’s rational introspective reflection to possess a certain temporal ordering, when it is not in fact genuinely so ordered (*ibid* 183)

The meta-experience notion makes it possible to make sense of this idea. For instance, there is no reason to think that when one experiences a touch on one’s nose and on one’s toe as simultaneous, the experiences of the two touch experiences must really be simultaneous. And it is unclear how “rational introspective reflection” would affect this impression of simultaneity.[[14]](#footnote-14)

The Exportation Principle stated above is thus not accurate. It should be amended as follows:

Exportation of Temporal Operators (Amended)

A subject perceptually experiences events E1 and E2 as standing in temporal relation R if and only if her perceptual experiences of E1 and E2 **are sensed as** standing in temporal relation R, subject to the proviso that she perceptually experiences E1, E2, and the temporal relation R.

Temporal order is not passively received. The experience of temporal order does not simply arise from the temporal order of the experienced events. Rather, it results from a process that bears at least some similarity to the one that Kant envisaged. The process is, as he urged, cognitively managed—I would prefer to say “managed by the perceptual system.” It is not driven by information that impinges on the senses. Rather, it is an *a priori* template for the ordering of perceived events.

How do the Kantian theses of Commonality, Aprioricity, and No Privileged Modality fare when applied to this account of our experience of time?

Commonality is relatively obvious: the song is heard; the dance is seen; they are sensed as temporally matched. This matching cannot be specific to vision or audition.

Aprioricity is also fairly straightforward, at least from the point of view of psychology. The discussion leading up to the Amended version of our Exportation Principle indicates that the perceptual system has a large role in the temporal measure of the events that we apprehend. Whatever this role is exactly, it is unlikely that it could be learned. It is not that temporal measuring is fully in place at birth. In the months after birth, infant perceptual systems are tuned up so that their representation of temporal relations becomes more accurate. But this cannot be a learning process in any straightforward sense because the system can never have access to temporal relations independently of its own timing mechanisms. It cannot, for example, *learn* that neural signals originating in the toe take a longer time to arrive at the brain than signals that originate in the eye, because it has no access to an independent clock that would time these intervals. Thus, it does not have the materials with which to learn how to correct for errors in its own timing by reference to the external world; all it can do is fiddle with the timing until it works.[[15]](#footnote-15) Either the system has either to have a built-in mechanism that regulates timing, or, assuming that temporal measuring is acquired over a period of time, it has to have a tuneable mechanism that arrives at an accurate measure by feedback and correction. It is likely that any such mechanism would operate on synchronization between processes, so that they work harmoniously together.

Pre-Modality seems to follow from a simple but pervasive characteristic of perceptual experience, namely that such experience presents its objects as happening *now*, i.e., as simultaneous with the experience itself. As a consequence, the temporal order of perceptual experience is projected onto the temporal order of perceived events. Thus, if a *perceptual* experience of event *E1* seems earlier than a perceptual experience of event *E2*, then since each of these perceptual experiences presents its object as occurring simultaneously with itself, it follows that together with its object experiences, the meta-experience presents *E1* as occurring earlier than *E2*. (Notice that this is a characteristic of perceptual experiences alone. An iconic recollection of *E1* could be followed by one of *E2* but this would not imply that *E1* was being presented as occurring earlier than *E2*.) This gives substance to Kant’s claim that time is the form of *inner* intuition. We feel the order of perceived events derivatively, by means of the temporal ordering of perceptual experience. But this entails that the temporal frame is pre-modal: it is an ordering not of visual or auditory events *per se*, but an ordering of *experiences* of visual and auditory events. This ordering is *non-*modal because of Commonality. It provides a framework for the ordering of perceived events.

Obviously, the analogue of the *now* condition fails for experience of space. Not everything is presented as occurring *here*, i.e., where the experience is. One does not, for instance, experience two things *X* and *Y* as ten feet apart by having an experience of *X* that is apparently ten feet away from the experience of *Y*. (This would happen only if one were to move ten feet every time one had such an experience.) There is, thus, a crucial disanalogy between space and time: the spatial relations of things are not projected from a modality-neutral proxy. This disanalogy vitiates any attempt to show that the representation of *space* is pre-modal in exactly the same way as that of time.

A different kind of argument is needed to support the Pre-Modality Thesis for space. This is what I shall now try to provide. As we shall see, Pre-Modality for the perceptual representation of space is somewhat analogous to the case of time. The analogy is routed through the Commonality thesis: just as time is a common measure of events presented by different modalities, so also space is a common measure as well. As we shall see, the most plausible account of the common measure entails Pre-Modality.

### Active Multimodal Perception

*Active perception* is purposeful activity in which a subject investigates the perceptual properties of a thing by interacting with it. For example, she might test the sharpness of a knife by gingerly running her thumb across it: this requires the coordination of touch and bodily motion and, for most of us, of vision as well. Or she might investigate the regularity of a shiny surface by feeling it while also looking for how light reflects off it as she manipulates it in her hands. Again, she might locate a squeak in her couch by bouncing on it in different places and listening for where the squeak seems to come from. In each of these cases, the subject investigates a thing by acting on it and observing the results; she knows what she does by monitoring her own activity; she uses the information that she gets from different modalities to determine exactly how the thing is. Clearly, this requires a coordination of spatial representations in various modalities, as well as those that guide action.

Or consider audition. It receives two sonic images, one from each ear. Each of these images is a frequency-amplitude function: that is, the basilar membrane in each ear separates the incoming auditory signal into frequencies, and measures the amplitude of each frequency as it occurs in the signal. From these two images, the auditory system constructs an image of sound sources distributed in three-dimensional space. When you listen to a symphony orchestra and a new instrument enters—a bassoon, say—the auditory system manages to detect a coordinated group of notes with similar timbre and separates it out of the soundscape as a separate sound. This permits you to locate it accurately enough to enable you to pick the bassoonist out visually: perhaps his body is moving in synchrony with the bassoon melody you hear. Once you single out the bassoonist, the sound seems to be more precisely located. In this sort of case, the auditory and visual object-location systems seem to be acting together and reinforcing each other in a temporally extended process.

For the sake of discussion, here is an account of a fairly complex piece of active multimodal perception:

One day a while ago, I felt a tiny stab between my shoulder blades. I couldn’t tell what it was: it could have been an irritation of the skin, or perhaps something poking at me. It took no particular cleverness to find out, and what I did next is exactly what most would have done. I squirmed and wiggled my shoulders. At this point, it became obvious that there was something sharp poking and rubbing against my skin, for I felt it move across my skin as I wiggled my shoulders. Moreover, since the movement of the sharp point seemed to coordinate with the movement of my jacket across my back, it was clear that there was something sharp caught in the cloth. I took the jacket off, but could see nothing. But grasping and bracing the cloth with my fingers and feeling around with my thumbs, I finally detected what was wrong. A stiff plastic thread had come loose from the padding of the jacket, and its sharp end was poking through the lining. Looking closely, I was finally able to see it. It was thin and more or less transparent. No wonder I had not been able to see it until then.

Suppose I had looked at the outside of the jacket and seen a similar thread poking through. It would have been natural for me to wonder whether it was the same thread. What would I have done to find out? Perhaps I would have tried to feel along its length. Or I could have pulled the thread on one side to detect movement or tension on the other.

These ways of poking and pushing at the world are low-tech. We often see animals undertake actions like this. Suppose you glue your dog’s toy to the ground. He goes to play with it, and finds that it doesn’t move. Or suppose you put it into a fretted box with a concealed entrance. He can see the toy but doesn’t quite know how to get it. Faced with these unusual situations, dogs will nose and poke and pull and play around with the toy in ways that are similar to the actions by which I got at what turned out to be a thread in my jacket. Similarly, for cats and birds, and (of course) primates. Perceptual skills can be honed and refined; they can even be taught and learned. But they are available to anybody—indeed any higher organism—with normal sensory and motor capacities.[[16]](#footnote-16)

### Scenes: Isotropic Perceptual Models

When I investigated the thread in my jacket, I was able to construct a complex multisensory model, consisting of a spatial configuration of perceived things and their sensory properties. This model is perspective-independent, or *isotropic*: even when I move my jacket, I can see and feel that the thread retains its location relative to other parts of the jacket. Very briefly, such a model, which I will call a ***scene***, is a three-dimensional configuration of things. Formally, it consists of at least the following:

1. A set of perceivable things, *P*;
2. for each member, *x*, of *P* a set of sensory qualities instantiated by *x*; and
3. for each pair {*x*, *y*} of members of *P*, the distance between *x* and *y*.

Isotropic perceptual models, or scenes, are evident in iconic memory.[[17]](#footnote-17) Think of a house in which you spent many childhood years. You know how the front hallway looked from both ends, how conversation in the living room sounded when you were standing in the kitchen and how it grew louder and more distinct as you walked down the hall toward the living room, the tactile feel of the bannister and how your hand got hotter as you rubbed against it on the way down, and so on. You can, at will, generate the perceptual experience of skirting the house clockwise, or counter-clockwise; you know how it was to come in from the back, and from the front. Each of these memories is experienced from a perspective; that is, the mental view of the house when you approach it from the front, is a view from the front. It is not, as it were, a “view from nowhere.” But each can be generated at will, including some that you never actually experienced yourself. It’s possible that you never climbed out on the roof above the porch, but you can easily imagine what the house looks like from there. (We will encounter similar considerations concerning isotropic scenes in the following section, when we consider the ideas of shape entertained by a blind person.)

Perspectival views are generated from an isotropic model. The alternative is to suppose that the mind stores a vast bank of perspective dependent images, which it draws on in the right order to string together a mental pathway. This alternative is implausible for several reasons. Consider first the free alternation between “field” and “observer” perspectives in iconic memory. When you mentally recall walking to the front door of your house, you can generally do it in two modes. You can experience the event from your own perspective, i.e., as if you were looking out of your own eyes, so to speak. This is the field perspective. Or you can experience it from outside your own body, viewing yourself as a part of the scene. This is the observer perspective. Since the observer perspective is (of course) not one that you could ever have experienced, it is clear that *this* perspectival view is generated from some underlying representation. And a similar thing happens when you generate images that you have never experienced, for instance, the view from the roof above the porch mentioned in the preceding paragraph. This shows that there is a mental process by which views from different perspective are generated.

A second point to consider is that isotropic scenes correspond quite closely to what Shepard and Metzler (1971) and Kosslyn (1980) call mental images (though they were concerned exclusively with vision). These psychologists showed that we are able mentally to manipulate mental images: we can rotate them, zoom in and out, change some aspects, and so on. For instance, imagine an unsymmetrical shape, such as a Swiss Army knife with several of its attachments extended—say a scissors pointing leftward, a bottle opener extended upward, a knife blade downward. One can mentally recreate how this would look if rotated: where the three instruments would point, how the foreshortening would change, how the protrusions would recede and approach relative to one another, and so on. Like the alternation between field and observer perspectives, this indicates that the underlying mental images do not *themselves* encode observer perspectives. They are stored in memory; and perspective-dependent views are constructed from them on the fly. But they are always presented from a perspective when they are entertained. When these scenes are mentally entertained, they yield what Christopher Peacocke (1983) has called a *scenario*, which consists of a point-of-view, and featured things with sensory qualities at locations specified relative to this point of view.(Perceivable *things* are an addition to scenarios as Peacocke envisaged them.)

♫♬♪

Rough face

Heavy and shiny

Scene with sample elements: The pyramid is an object; and so are its significant parts—including faces, edges, and so on. These objects have features: shape, colour, sound emission, etc. Since the distances between the objects are constant, the whole scene is a rigid solid that remains the same under rotation, and as the perceiver approaches or recedes.

I have said that when an isotropic scene is observed from different perspectives, we are presented with different scenarios. However, I would contend that even in situations where perception gives us no more information than is contained in a single Peacocke-style scenario—as would be the case in a single short-duration glimpse of a situation—perception is always influenced by perspective-independent scenes, or is directed toward constructing them. With every movement and every shift of attention, we continually and always undertake a process of building this scenario up into a scene. When, on the other hand, we perceive a familiar thing, we experience a scenario that is pregnant with the scene (or scene fragment) that we have already built up: for though we only glimpse the thing from one point of view, what we see is influenced by this.

Summing this up, we have:

*Isotropic Multimodality Thesis* Retained perceptual images are (to varying degrees) independent of the subject’s perspective or point of view. When a mental image is episodically entertained, it is always from a perspective. This perspective-dependent image is generated from the retained isotropic model. Both retained and episodically entertained images are multimodal. [[18]](#footnote-18)

These are somewhat dogmatic claims, and I will not attempt to justify them any further here. My concern is with the role of the representation of space in active perception. Constructing an isotropic scene—a scene that retains its spatial encoding under rotation—requires that our perceptual systems collate features provided by different modalities in locations common to those modalities. In short, it requires a cross-modal representation of space that specifies locations in an observer-independent coordinate system.

### Three Grades of Multimodal Involvement

The fact that more than one sense can be relevant to action has not, of course, escaped the attention of philosophers and psychologists. In fact, thanks to Molyneux, it was one of the central problems of philosophy in the 18th century. How are the spatial perceptions, or sensations, of different modalities related to one another? Approaches to this problem vary in their commitment to real integration, as I shall now recount.

1. The Privileged Modality View: Cross-Modal Matching

The traditionally most accepted idea about the multimodal representations of space is that there are none. Any given property, spatial or otherwise, is represented in just one modality, and the other modalities do not represent it, though they may represent qualities associated with it.

Berkeley’s treatment of the Moon illusion is a good illustration of this. In the illusion, the “confused” or shimmery-blurry character of the Moon on the horizon is associated with greater distance. In other words, the Moon appears further away when it is on the horizon than at its zenith because it looks more confused. Because the same-sized retinal image indicates greater size given greater distance, it follows that the Moon appears larger when on the horizon. *Distance* is not, however, visually represented, according to Berkeley (*New Theory* §2-20); for the way a point is projected onto the retina is independent of its distance. Distance is represented by touch[[19]](#footnote-19)—by the sensation of reaching for a thing, or travelling towards it, or by turning one’s eyes inward when looking at it. Distance is a tactile idea that is simply associated with visual confusedness. Similarly, size is not visually represented: the size of the retinal image is co-determined by the size of the projecting object and its distance. Size too is a tactile quality inferred from distance, which in turn is inferred from the visual appearance of confusedness.

One problem with the Privileged Modality view is that there are reasons to assign some spatial properties to one modality, and some to another. Berkeley thinks that *distance* and *size* are represented tactually, not visually. However when it comes to *shape*, some think that vision has an advantage. (See Stokes and Biggs, this volume.)

Diderot offered one canonical argument to the effect that touch does not represent shape. According to him, there is *no* simultaneous tactual idea of shape.

How does a man congenitally blind form ideas of shape? I believe that the notion of direction is given to him by the movements of his body, the successive existence of his hand in different places . . . If he runs his fingers along a taut thread, he will get the idea of a straight line; if he follows the line of a sagging wire, he will get that of a curve . . . For a blind man, unless he be a geometer, a straight line is nothing but the memory of a succession of sensations encountered along a taut string . . . Whereas we combine coloured points, he combines only palpable ones, or, to speak more exactly, only his tactile memories. (“Letter on the Blind,” translated in Morgan 1977, 39)

A blind person takes in shape by a temporally extended process of feeling a thing, Diderot suggests, while a sighted one takes it in in a single, temporally punctate, visual act. But *shape* is nota temporal succession. And in any case it is not any particular succession that a blind person might employ on a given occasion—feeling the broad end of an arrowhead first and the point later is a succession of sensations different from feeling the point first, but the shape is one and the same and hence not identical with either. The blind man’s experience is thus not of *shape* as such. Let’s call this Diderot’s Thesis: Diderot thought that shape is a visual idea that the blind represent abstractly.

It makes no sense to think that spatial properties can be distributed across modalities. For the spatial properties are defined and measured by space and the question arises: to what modality does the representation of space belong, and how can a modality-specific representation of space be applied across modalities, as well as to “abstract” quantities? The spatial properties, such as size, distance and shape, are an inter-related group, and it makes little sense to split them up. The Privileged Modality view is in trouble if it is unable to give a unified treatment of all spatial properties as belonging to the same modality.

Another, and in my opinion, a decisive objection to the Privileged Modality view is simple introspection. Berkeley holds that we have no visual idea of size. But this is simply false: things almost always *look* as if they are a certain size. (Admittedly, very distant things can often look indeterminately large, but this is irrelevant to my point.) It is simply obvious that size is represented visually as well tactually.

Gareth Evans (1985) notes that it is absurd to think that direction is represented *indirectly* by some modalities:

No one hears a sound *as coming from the same side as the hand he writes with* in the sense that, having heard it thus, he has to say to himself “Now I write with this hand” (wiggling his right hand) so the sound must be coming from there (pointing with his right hand)” (384)

The same sort of thing can be said of size. One doesn’t see something as large by firstforming the idea that one can touch it by stretching out one’s arm, and then noticing that large things project an image of this angular size when they are an arm’s length away. There is a simple visual impression of size. This is precisely what the Moon Illusion illustrates. The Moon *looks* large on the horizon. It does not simply look confused and of the same angular size as when it is at its zenith.

Associationism does not make things easier for Berkeley. His idea is that size is inferred from distance, which is in turn inferred from visual confusedness. But there is something in the phenomenology that this explanation passes over. To account for the phenomenology of visual appearance of size, Berkeley must account for how things come to *look* a certain size. But he cannot do this. *Seen* size does not exist for him; his theory simply denies that anything *looks* large or small. Whether cross-modal matches are made innately or by past association is not the central sticking point for him. It is the *look of size* that his Privileged Modality view cannot accommodate.

The driving idea behind the Privileged Modality view is that cross-modal correspondences are post-perceptual. Since perception is a process that begins and ends within a single modality, the appearance of sameness between spatial relations as seen and spatial relations as felt cannot be ascribed to perception. Of course, there is no reason, as Evans remarked, why there should not be an appearance of similarity between a spatial relation or shape as seen and as felt[[20]](#footnote-20):

There is nothing on the most radical empiricist view that precludes sensations produced by the stimulation of different sense modalities being sufficiently close together in the innate similarity space for responses conditioned to the one to generalize to the other. There is nothing particularly upsetting to an empiricist theory of concept formation in the suggestion that human subjects who are trained with the use of ‘harmonious’ in the case of sounds might generalize its use (without further training) to the case of certain combinations of colours. (376-377)

However, as Evans recognizes (by his talk of “concept formation” above), such an “appearance” of similarity cannot be perceptual in origin since perception does not reach into any conscious state in which cross-modal comparisons are made. The similarity of felt and seen shape is thus of a very different sort than that of red and pink, or any other two visual qualities.

1. The Behavioural Space View: Multimodal Calibration

Let us say that something looks to be over to the left of me. It cannot look this way *merely* in virtue of a certain visual sensation with modality-specific phenomenal character that marks things as over on the left—call it “visual leftishness.” For, as Evans forcefully argues (following George Pitcher 1971 and Charles Taylor 1979), it would then be possible to sense a perceived thing visually leftish, but fail to know that one had to stretch out one’s left arm to reach for it. Evans writes:

We do not hear a sound as coming from a certain direction, and then have to *think* or *calculate* which way to turn our heads to look for the source of the sound . . . Since this does not appear to make sense, we must say that having the perceptual information at least partly consists in being disposed to do certain things . . . The subject . . . hears the sound *up*, or *down*, *to the right* or *to the left* . . . It is clear that these terms are *egocentric* terms; they involve the specification of the position of the sound in relation to the observer’s own body. (Evans 1985, 382-383)

Visual and auditory sensations of location are not contingently connected to location in behavioural space. Evans concludes that in all modalities, these locations are “specified in the same, egocentric, terms” and are “used to build up a unitary picture of the world.” “There is only one behavioural space,” he says (390), meaning that in every modality, space is specified in behavioural terms.

This is an extremely important argument. Visual leftishness is necessarily connected to the feeling that one would have to stretch out one’s left hand, or walk leftwards, to reach for it. Equally, *auditory* leftishness is necessarily connected to the same behaviour. Spatial awareness is coordinated across all of the senses by coding that is appropriate to the motor system. And this seems to hold up empirically. Pawan Sinha’s investigation of newly sighted individuals in India (Held *et al* 2011) reveals that it takes these patients a few hours, at the least, to correlate shapes across the modalities. But he and his colleagues do *not* report that these subjects experience any difficulty at all with visually guided reaching or navigation (as surely they would have if they had noticed any such difficulties). They are not, for example, reported as deficient with regard to looking over to the left when they hear a sound from there, or with regard to pointing or turning toward a sudden bright flash of light.

Evans’s argument runs parallel to the argument about the representation of time in section II above. It too is what one might call a “Common Measure” argument. The argument of section II established that time must be pre-modal because, across all modalities, experiences have a common temporal dimension. Evans’s argument is also a common measure argument. The idea is that since one can deliberately move one’s own body relative to anything that one experiences through the senses, so anything that is experienced through the senses must be experienced as located relative to one’s own body. Perception, taken as a whole, offers us, as Evans says, a “unitary picture” of things laid out in space.

As with time, it is important to recognize that the common measure is given through a *representation* of spatial properties. As with time, this placement of things in a common perceived space is not *automatic*: it is the result of perceptual data processing. And, as noted in section I, bodily sensations may possess location in a scheme that does not smoothly coordinate with that of the external modalities. On occasion, the perceptual system gets it wrong: this is illustrated by the “ventriloquist illusion” in which the system locates the ventriloquist’s voice in the moving mouth of the dummy. Again, there is the “rubber hand illusion” in which a subject’s hidden hand and a visible rubber hand are simultaneously stroked—subjects report that they feel the stroking where the rubber hand is (Botvinick and Cohen 1998). As the experimenters observe, “this illusion involves a constraint-satisfaction process operating between vision, touch and proprioception—a process structured by correlation normally holding among these modalities.” In other words, the system’s activity in placing modal stimuli in a common space results, in this case, in an illusion.

The Common Measure argument regarding space is, as I said earlier, extremely important. However, the Behavioural Space view suffers from some of the same difficulties as the Privileged Modality view. Consider first the idea that visual experience of location has a characteristic phenomenal character. Some things *look* as if they are over on the left. This phenomenal character that marks the look of being on the left is not the same as the phenomenal character that marks a sound of being on the left. Yet, if visual and auditory space were simply behaviourally specified, these would both amount to awareness of the behaviourally specified location of such things—and thus be the same. But neither the look nor the sound is, as such, the awareness of where one would have to reach for it. For something to look or sound as if it were over on the left is not *the same* as for it to appear as if one would have to reach for it over on the left. The behavioural awareness *comes with* and is *integrated with* the look and with the sound, which are in turn integrated with one another in this respect. But they are not all the same act of awareness.

Secondly, and quite crucially, the Behavioural Space view falls prey to the counter-intuitiveness of Diderot’s Thesis. For a blind person, Diderot says, any given encounter with a three-dimensional shape other than those small and simple enough to register all at once on the surface of the skin, is temporally extended. (One can imagine that a small circle or globe might register in this way, but what about a small pyramid? The simultaneous feel of the four angles of a small pyramid is clearly insufficient to determine the nature of the solid.) Moreover, there are many ordered successions of experiences that a blind person can undergo while tactually exploring any given shape. Thus, if we are to attribute a simultaneous idea of shape to a blind person, it must be an idea that abstracts away from any succession of tactile experiences in much the same way as an isotropic scene abstracts away from any particular perspective-dependent view. This analysis is confirmed by the testimony of the blind scholar, Pierre Villey, on his mental images of three-dimensional things. Villey concedes that his tactual perception of such things is sequential, but:

. . . if, an hour after feeling it, I search in my consciousness for the memory of the vanished chair . . . I do not reconstruct it by means of fragmentary and successive images. It appears immediately and as a whole in its essential parts . . . There is no procession, even rapid, of representations . . . I couldn’t tell in what order the parts were perceived by me. (Evans 1985, 369, quoting Villey 1922, 183-184)

Villey’s account was that his tactual sequences were smoothly converted into “simultaneous” representations.

But this raises a question about vision. Clearly, we do not visually grasp large and complex shapes in a single visual act. We have to scan them. But one person may scan a large figure left to right; another may scan it right to left. Clearly this does not mean that each will arrive at a different kind of grasp of tis shape. Now, when it comes to three-dimensional shapes, it is simply impossible to visually take them in in a single act. What, then, is the nature of our visual grasp of a cube or a sphere if Diderot is right? Following his reasoning, we should conclude that it abstracts away from perspectival experience in the same way as a blind man’s does. But this means that we grasp visual shape by transcending the egocentrically specified views of momentary experience. But this is precisely what I argued in section IV above. Retained multimodal images discard at least some of the egocentricity of immediate experience, and this demands a representation of space that is cross-modally commensurate.[[21]](#footnote-21)

Evans (1985) thought that the counter-intuitiveness of Diderot’s Thesis was the central difficulty for a negative answer to Molyneux’s Question. Properly applied, Diderot’s reasoning indicates that there cannot be, in *any* modality, a “simultaneous” idea of three-dimensional shapes. Any simultaneous idea of a three-dimensional shape must be amodal at least in the sense that it is not the same as the content of any modality-specific experience, or succession of experiences. It is ironic that Evans, who argued strongly against a view that makes haptic images of three-dimensional things sequential, ended up adopting a view that makes all images of things, including visual images, sequential.

1. The Pre-Modal Space View: Multimodal Integration

Active perception is an activity that involves the coordinated use of all the different senses. And it enables us, in an instinctive and largely untutored way, to arrive at more or less isotropic models of the world outside the mind. The coordination of both process and product argue for a level of integration among modality-specific representations of space that the Privileged Modality view cannot accommodate. The Behavioural Space view attempts to fill this lacuna by making the representation of space extra-modal. However, the egocentricity of perceptual representations of space that this view implies results in a fragmentation and temporalization of perceptual models. The only way to accommodate isotropic multimodal images is to posit a pre-modal representation of space.

I will conclude now with a brief discussion of some main elements of a Pre-Modal view of the perceptual representation of space.

### The Pre-Modal View

Think of an old car. You are looking at it as your friend starts it up. When the engine comes to life, you hear the rumbling noise; you can see and hear the hood vibrate; when you touch it, you can feel the vibration; as the engine warms up, you pick up the faint odour of motor oil. All of these events are sensed as localized in the same region: the sound is perceived as coming from the car, even the smell, if you move around and sniff. (Think of a garage full of cars: the enclosed space smells of oil, but you are told that only one of the cars is leaking oil. Can you find it by moving around and sniffing?)

In the pre-modal view, there is a representation of space that underlies the location of these sensory qualities: a three-dimensional matrix that is able to receive features. (Time also figures in this, but I shall leave it out for present purposes.) As each sense provides the active perception system with information, features are pasted onto locations in this matrix. To a static observer using only one sense, the locations are egocentrically identified. If, for example, you are looking at something from a stationary perspective, you will have information about how visual features are distributed in space, relative to your eyes. The use of the other senses, even from this fixed observation post, provides information that is somewhat more perspective-independent. Touch, for example, provides information relative to parts of the body other than the eyes. Haptic features such as warmth or vibration are felt by the hands, but are nevertheless referred to the thing that you see yourself touching. (Think of the rubber hand illusion: the feeling of being stroked by a brush is referred to the rubber hand that you see, not to your own hand, which is being stroked and is the origin of the sensation.) Movement increases the perspectival independence of the perceptual model. You discover previously unsensed features, which are then incorporated into your mental image. In the limit, your perceptual model is of a rigid solid replete with sense-features. The underlying representation of space enables the construction of this model.

The pre-modal view of space can be likened to the representation of space in a touch screen device. You touch a visual icon on your smartphone and it reacts appropriately. It does not do this by cross-identifying the representation of space that governs its icon placement with the representation that locates your pressure on a part of the screen. Rather it has a single underlying representation of space relative to which the touch and the icon are both located. When it detects your touch, it looks up what icon occupies that same location. Multimodal scenes are constructed in a similarly pre-modal manner.

There are two sources for the construction of an isotropic scene: the features detected by each sense and the spatial matrix in which they are placed. As in the case of smartphone touchscreen, the latter is not provided by the senses. Rather, as Kant realized, it is a pre-modal framework needed for the spatial coding of perceived scenes. Of course, the senses provide enough spatial information to determine, for example, whether one thing is to the left or the right of another. But the pre-existing model regiments such information. For instance, although auditory spatial information is much coarser in grain than visual information, sounds are precisely located in things that appear to be emitting them. (For instance, vaguely located voices are precisely located in a moving mouth, as in the Ventriloquist Illusion.) And where touch and vision give information about the same thing, the information is reconciled, giving due weight to the reliability of each in the situation being examined. (See, for example, Ernst and Banks, 2002; Millar and Al Attar, 2005). Perceived spatial relations will, therefore, have an *a posteriori* component. Kant’s insight, however, was that every spatial relation has an *a priori* component that depends on the structure of the underlying matrix.

The pre-modal view faces a difficulty that we cannot fully resolve here—the apparent incongruity of spatial relations in bodily sense and the external senses. As mentioned earlier, I may feel an itch in my finger and observe visually that my finger is in front of the TV without feeling the itch in front of the TV. Two observations will, however, help to suggest a fruitful approach to this question. The first observation is that although the bodily sense locates feelings in a body scheme—that is, these feelings are felt to be located in parts of the body—it is nevertheless the case that subjects have a sense of how parts of the body are located relative to external things. The second observation is that there is a certain incoherence of spatiality within the body sense itself. For instance, as Ned Block (1983) has observed, I may feel a pain in my thumb and feel that my thumb is in my mouth without feeling the pain in my mouth. The second observation suggests that the body sense has a special and separate sense of space, one that is located *only* relative to body parts, and which does not carry information about the location of these body parts relative to other things (including other body parts). The first observation suggests that though the special spatial framework of the body sense is not integrated into the pre-modal representation of space, the bodily sense also participates in pre-modal sense. For instance, it is possible to sense whether the index fingertip of your raised right hand is higher or lower than your nose. Thus, it might be that feelings such as itches and pains are localized only relative to body parts, but that body parts themselves are localized relative to the external objects of vision, etc. However this may be, the safest course here is to allow that the bodily spatial sense is somewhat anomalous and may at best partially and imperfectly participate in the pre-modal representation of space.

*Conclusion* The traditional notion is that we perceive space is through sensory information. It was, and still is, thought, for example, that spatial information is contained in the biretinal and the binaural images. This conception of spatial perception is unable to account for the construction of isotropic multimodal scenes in active perception. It is often said that active perception uses cognitive resources far in excess of the sensory process alone. And this is undoubtedly true, even on a very capacious view of the sensory process. Nevertheless, the system must coordinate the spatial images that each modality provides in order both to control action and to construct isotropic scenes—each modality-specific sensory image needs to be coded in a way that allows it to be collated with every other. The Privileged Modality view and the Behavioural Space view offer inadequate accounts of this process. What is needed is much more like the coding used in a touchscreen: a pre-existent spatial representation that provides a coordinate space for the placement of features from all of the modalities. Kant’s view that space and time are *a priori* intuitions conceives of space and time in much this way—i.e., as common measures of information received by the different modalities. The perceptual representation of space is not, as he clearly saw, proprietary to any one modality, and it is not behavioural space.

## REFERENCES

Block, Ned (1983) “Mental pictures and cognitive science,” *Philosophical Review* 92: 499- 541.

Botvinick, Matthew and Cohen, Jonathan (1998) “Rubber Hands ‘Feel’ Touch That Eyes See,” *Nature* 391: 756.

Eagleman, David M. (2009) “Brain Time,” in M. Brockman (ed.) *What's Next? Dispatches on the Future of Science: Original Essays from a New Generation of Scientists* (New York: Vintage): 155-169.

Eagleman, David M. (2010) “How Does the Timing of Neural Signals Map Onto the Timing of Perception?” in R. Nijhawan and B. Khurana *Space and Time in Perception and Action* Cambridge: Cambridge University Press: 216-230.

Ernst, Marc O. and Martin S. Banks (2002) “Humans Integrate Visual and Haptic Shape Information in Statistically Optimal Fashion,” *Nature* 415: 429-433.

Evans, Gareth (1985) “Molyneux’s Question,” in A. Phillips (ed.) Gareth Evans *Collected Papers* Oxford: Clarendon Press: 364-399.

Held, R., Ostrovsky, Y., deGelder, B., Gandhi, T., Ganesh, S., Mathur, U. and Sinha, P. (2011). “Newly sighted cannot match seen with felt,” *Nature Neuroscience*14: 551-553.

Hopkins, Robert (2005) “Molyneux’s Question,” *Canadian Journal of Philosophy* 35: 441-464.

Husserl, Edmund (1905/1964), *The Phenomenology of Internal Time-Consciousness*, M. Heidegger (ed.) and J. S. Churchill (trans.). Bloomington: Indiana University Press.

James, William (1890) *The Principles of Psychology* New York: H. Holt.

Kappers, A.M. and Koenderink, J.J. (1999) “Haptic Perception of Spatial Relations,” *Perception* 28: 781-796.

Kosslyn, S.M. (1980) *Image and Mind*. Cambridge, MA: Harvard University Press.

Matthen, Mohan (2010a) “The Sensory Representation of Color,” inJ. Cohen and M. Matthen (eds.) *Color Science and Color Ontology* (Cambridge MA: MIT Press, 2010): 67-90.

Matthen, Mohan (2010b) “Is Memory Preservation?” *Philosophical Studies* 148: 3-14.

Matthen, Mohan (forthcoming) “How To Be Certain: Sensory Exploration and Empirical Certainty,” *Philosophy and Phenomenological Research* (published online, 2012: DOI: 10.1111/j.1933-1592.2011.00548.x)

Millar, Susanna and Al Attar, Zainab (2005) “What Aspects of Vision Facilitate Haptic Processing?” *Brain and Cognition* 59: 258-268.

Millar, Susanna (2008) *Space and Sense* Hove, England: The Psychology Press.

Morgan, Michael J. (1977) *Molyneux's Question: Vision, Touch, and the Philosophy of Perception* Cambridge: Cambridge University Press.

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

Phillips, Ian B. (2008) “Perceiving Temporal Properties,” *European Journal of Philosophy* 18: 176-202.

Phillips, Ian B. (forthcoming) “The Temporal Structure of Experience,” in D. Lloyd and V. Arstila (eds.) *Subjective Time: the Philosophy, Psychology, and Neuroscience of Temporality* Cambridge MA: MIT Press.

Pitcher, George (1971) *A Theory of Perception* (Princeton: Princeton University Press).

Sassen, Brigitte (2004) “Kant on Molyneux’s Problem,” *British Journal for the History of Philosophy* 12: 471-485.

Shepard, R. N. and Metzler, J. (1971) “Mental Rotation of Three-Dimensional Objects,” *Science* 171: 701-703.

Siegel, Susanna (2006) “Subject and Object in the Contents of Visual Experience,” *Philosophical Review* 115: 355-388.

Taylor, Charles (1979) “The Validity of Transcendental Arguments,” 79: 151-165.

Villey, Pierre (1922) *The World of the Blind* London: Simpkin, Marshall, Hamilton, Kent & Co.

Wittman, Marc (2009) “The Inner Experience of Time,” *Proceedings of the Royal Society B* 364: 1955-1967.`

1. *Pains* are felt to be spatially located in the body—and in a subject-perspective independent way, to boot (Matthen forthcoming). Kant ought, therefore, to ascribe to them the same kind of felt objectivity that he ascribes to visual stimuli: that is, he should hold that pains are bodily disturbances, not “feelings” or inner experiences. By contrast, thoughts and emotions are not felt as located in the body (or anywhere else). He should count them, therefore, as mental occurrences apprehensible only by outer intuition. [↑](#footnote-ref-1)
2. Actually, this is not exactly correct. Susanna Siegel (2006) has argued that in order to appear as if they exist outside the mind, things must look as if their location does not depend on the subject’s location and perspective. Floaters and phosphenes are apparently spatial, and are spatially related to each other, and so they satisfy Kant’s requirement. However, they fail Siegel’s test; they are always in the same place *relative to the subject*, so their apparent location in objective space changes as she moves. And they are not seen as existing independently. But even the subject-independent location test doesn’t quite do the trick. For though Siegel’s thesis is plausible concerning vision, it is less so with regard to audition. Music heard through earphones has subject-dependent location—it comes from the left or from the right of the subject even when she moves—but it appears to exist outside the subject’s mind. [↑](#footnote-ref-2)
3. I am not entirely certain that the passage that I quote above is meant to apply to the Pre-Modality thesis. See, however, Gareth Evans (1985), esp. 369-370, which suggests that it is. [↑](#footnote-ref-3)
4. See Millar and Al Attar (2005). Brigitte Sassen (2004) argues that Kant does not require this kind of cross-identification, and that this is why he fails to mention Molyneux’s problem. All that he requires, according to her, is that each sense place objects “side by side,” and not that locations need to be “isomorphic” across distinct senses. I think she under-estimates the strength of Kant’s demands on cross-modal identification. First, if it visually appears that *X* is to the left of *Y*, then it should also haptically appear that X is to the left of Y. Secondly, the “side-by-side” relations in each sensory space should add up to a shared Euclidean structure. [↑](#footnote-ref-4)
5. Millar (2008) contains detailed synoptic discussion of empirical evidence regarding cross-modal matching. [↑](#footnote-ref-5)
6. It is unclear, but I think irrelevant, how Kant would have reacted to non-Euclidean geometry. I take his argument in the *Transcendental Aesthetic* to be about certain norms regulating the perceptual representation of space. It is not about “space itself”—which does not exist in Kant’s way of thinking. I take it that faced with Einstein’s theory that gravitation bends space, Kant *could* have conceded that physics requires a non-Euclidean space, and perhaps even that we could have a non-Euclidean *concept* of space, while denying that we can perceive space as non-Euclidean. [↑](#footnote-ref-6)
7. Perhaps there could be Escher-like presentations to the contrary, but it must be remembered that these are two-dimensional renderings of three-dimensional impossibilities. This makes a difference: he did not present us with three-dimensional violations of three-dimensional Euclidean geometry; nor do his drawings violate two-dimensional geometry. [↑](#footnote-ref-7)
8. For further discussion, see Matthen 2010a. [↑](#footnote-ref-8)
9. The surface reflectance view is over-simplified—luminances and transmittances are colours too. But this does not matter for my present point. [↑](#footnote-ref-9)
10. Not all perceptual coding effects give rise to apodeictic appearances. Weber’s Law states that the discriminability of perceptual magnitudes (such as brightness or volume) depends on the ratio of these magnitudes (not their absolute difference). (Thus, if a light of brightness 10 is just discriminable from one of brightness 11, then a light of brightness 20 will be just discriminable from one of brightness 22.) This is a coding effect. For as Stanislas Dehaene has written:

    Ernst Weber discovered what we now know as Weber’s Law: over a large dynamic range, and for many parameters, the threshold of discrimination between two stimuli increases linearly with stimulus intensity. Later, Gustav Fechner showed how Weber’s law could be accounted for by postulating that the external stimulus is scaled into a logarithmic internal representation of sensation. (Dehaene 2003, 145)

    There is, however, no apodeictic perception that corresponds to Weber’s Law, no feeling of certainty, no feeling of any sort, that the larger a stimulus, the more similar it appears to its close neighbours. [↑](#footnote-ref-10)
11. It has traditionally been supposed that since sensory experience is temporally punctate, duration cannot be sensed, and that somehow memory is involved in the experience of duration. I think that this is irrelevant: the iconic retention of experience in short-term memory and its temporal measurement is part and parcel of the perceptual process. [↑](#footnote-ref-11)
12. One point to be noted here is that one can experience a process *without* experiencing every constituent event. For example, I can experience an object moving from A to B *without* experiencing it located at every point in between. However, if I experience O moving from A to B without experiencing it at intermediate points C and D, then (I would contend) I do not experience a temporal relation between O being located at C and it’s being located at D. Thus, the example does not contradict the Exportation Principle. [↑](#footnote-ref-12)
13. There is a question about what admissible substituends for ‘*R*’ might be. There might be intuitive units of time, perhaps measured by internal clocks (Wittman 2009). If so, a person might experience an interval of time as thus-and-so long in this intuitive sense, and this interval of time might be the same as say 10s. Thus, to say that a person experienced two events as occurring 10s apart would mean that she experienced them as separated by an interval that happens to be the same as 10s, though the experience would not be sufficient to tell the subject that the interval was 10s. another point to note here is that the time-interval between *E*1 and *E2* must be short enough if time for time perception to operate. [↑](#footnote-ref-13)
14. On the other hand, Husserl (1905/1964, 31) seems to exaggerate when he writes: “Our ideas do not bear the slightest trace of temporal determinateness.” Husserl’s examples seem to counter the First Pass of the Exportation Principle: for example, “If, in the case of a succession of sounds, the earlier ones were to be preserved as they were while ever new ones were also to sound, we should have a number of sounds simultaneously in our imagination, but not succession” (32). Here it seems as if the sensations are simultaneous, and the notes are heard as simultaneous. [↑](#footnote-ref-14)
15. This is not to say that association and learning never plays a role: see the quote from Eagleman (2009) above. However, this is a case where the temporal relation between internal events gets established, not where the system is correcting for external inaccuracies to which it has not independent access. [↑](#footnote-ref-15)
16. For further discussion of active perception, with particular reference to its epistemological force, see Matthen (forthcoming). [↑](#footnote-ref-16)
17. For more about scenes in iconic memory, see Matthen 2010b. [↑](#footnote-ref-17)
18. Casey O’Callaghan (this volume) has an account of multimodal binding that is similar in intent, but somewhat different from this, inasmuch as it is event-based rather than object-based. Spence and Bayne (this volume) tentatively deny that there is simultaneous awareness of features in different modalities; they would not, however, deny the existence of isotropic multimodal models. [↑](#footnote-ref-18)
19. It would have been more appropriate to include proprioception, since both touch and proprioception are involved in estimating the distance one has walked or how far one is reaching toward an object. But Berkeley seems to adhere very much to the five-senses view. [↑](#footnote-ref-19)
20. John Mackie (1976) takes a similar line with regard to Locke. Since both the visual and tactile ideas of a cube resemble a cube, the newly sighted man should be able to visually differentiate a cube and a globe by the fact that the visual idea of the first resembles a globe, and that of the second a cube. This constitutes a relevant similarity between the tactual and visual idea of a globe. [↑](#footnote-ref-20)
21. An completely isotropic model includes all of the absolute distances between its members. A completely egocentric model specifies what sensory features lie in various directions relative to the perceiver. Some images lie in between these extremes. [↑](#footnote-ref-21)