# 

# 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 appearance, and time ***the***form of inner sense (B42, 49). Space is (b) “a condition of the possibility of appearance,” (B39), and (c) “found in us prior to any perception of an object” (B41). It is (d) “the subjective condition of sensibility, under which alone outer intuition is possible for us” (B42). Our representation of space is, moreover, (e) the source of our certainty regarding geometrical truths (B41).

Encapsulating the above in contemporary terms, we find in Kant the following:

*Pre-Modality Thesis*:

1. [*Commonality*] There is a single representation of space and of time common to all modalities. (a, c, d)
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, e)
3. [*No Privileged Modality*] No single modality has priority with regard to these templates. The common representation of space is not simply projected from vision or touch to the other modalities. (b, 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. (a, 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). Suppose that every sensory impression belongs to some modality. Sense impressions of different modalities differ in (what we would now call) phenomenal character. 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. (It is safe to say that this assumption retains much of its intuitive appeal today.) It follows that when there are sense impressions of the same quality in different modalities—for example, the visual and tactile impressions of sphericity—they would, nevertheless, differ with regard to their cognitive significance. Thus, it would be possible to *doubt* that they were of the same quality. This was Locke’s thought in response to Molyneux’s question. Kant swam against this tide. According to him, the sensory representation of space does not belong to any modality. This undercuts Locke’s approach because the representation of space would span the modalities. Kant’s thesis 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, and 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. 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 here 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. To me, it seems that Outerness has a good deal of merit and deserves further investigation, but it goes well beyond my aims here to try and construct a corrected version that would be worth defending.

### Kant’s Arguments

Kant’s Pre-Modality thesis has fallen out of view in philosophical discussions 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 perceiver’s mind because the senses locate them in space. In other words, he holds (in something of an echo of Descartes’ conception of matter as essentially extended) that the perception of things as spatially ordered is constitutive of perceiving them as existing objectively.[[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)

Hume’s position concerns *existence*.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 a perceiving subject “consider’d as” external, there could be an exactly similar impression that the subject takes as possessing noexternal correlate. Thus, he would have held, no sensory quality differentiates things perceived as outer and those perceived as inner. This is where Kant gets off the boat. 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 *constitutive of* 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. The further claim is that “one can represent only one space,” or, in other words, that one cannot sense two objects as outer without sensing that they are located in the same space.[[3]](#footnote-3) He claims, in other words, that there is only one representation of sensory space. In other words:

[1SS] To sense something as external is *eo ipso* to sense it as existing in the same space as everything else sensed as external.

1SS is, according to Kant, an *a priori* intuition. Clearly, his position is contestable. For one might think that the “one all-embracing space” is given to us by reason, and is not embedded in perceptual experience *a priori*. (I don’t agree with this objection to Kant, as will emerge later.)

Let’s grant Kant 1SS for the sake of argument. The question arises: Does 1SS 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.[[4]](#footnote-4)
4. [Aprioricity may be thought to follow from the thesis of 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 is exactly the point made by the different-phenomenal-character argument outlined 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). This, after all, is the problem posed by Molyneux.

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 pinprick does not feel as if it is in front of the TV. The TV, the pin, and the pinprick are all spatially located, one by vision, one by both vision and touch, and the last by touch. But perception 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.[[5]](#footnote-5) 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. My pre-modality view has to address these limitations. But at the very least, they show that one cannot argue for Pre-Modality *a priori*, as Kant seems to do.[[6]](#footnote-6)

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;[[7]](#footnote-7) these features of space are “apodeictic.” But nothing that we learn through the senses is apodeictic: any empirical proposition can 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 about the content of these cognitive necessities, and in particular about perceived space being Euclidean.) He argues further that only those propositions that are known by the senses are confined to a single modality. (See below.)

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, the 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. (This is the triangle inequality.) 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 *in virtue of appearing straight*, a line appears to be the shortest distance between two points.[[8]](#footnote-8) 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. And analogously to the triangle inequality, it seems necessary that pure *yellow* is more similar to *orange* than it is similar to pure *red*.[[9]](#footnote-9) Again, it is often suggested (by Wittgenstein, for example, in the *Remarks on Colour*)that the intuition that nothing could be reddish and greenish at the same time is not derived from colour experience *a posteriori*.[[10]](#footnote-10)The reddishness that presents itself as a component of *orange* also presents itself as impossible to combine with a greenish hue. 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.”[[11]](#footnote-11)

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 *must* stand in 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.[[12]](#footnote-12) 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.[[13]](#footnote-13)

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) Mental 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. Premise (iii) is the source of the error.

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

Our sense of time is clearly multimodal. 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 a flash of light and is jabbed on the finger with a pin, the flash and the jab are 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.)[[14]](#footnote-14)

All awareness is *inter alia* awareness of time. So it seems that time is a common measure of events sensed both within and across modalities. How does this work? And is the representation of space cross-modal in a similar way?

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 the temporal structure of experience represents the represents the temporal structure of experienced events. For example, you experience a song as taking a certain amount of time by your experience of the song taking that amount of time.

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, [[15]](#footnote-15) and the temporal relation R.[[16]](#footnote-16)

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*.

At first sight, the Exportation Principle offers 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). Similarly, a time lag of less than 150 ms between a subject’s finger movements and visual data regarding these movements is disregarded: when the lag is longer the seen movements are not perceived as the subject’s own; when it is shorter, they are (Hoover and Harris 2012).

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.[[17]](#footnote-17)

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

The extraction of temporal structure from the data stream holds for external events as well. The McGurk effect (see the Introduction to this volume) shows that speech perception depends in part on vision: the visual perception of the speaker’s mouth is relevant to what phoneme is heard. It is *also* true that the perceived time of the voice stream depends on vision. For even when the auditory stream lags the visual stream by as much as 170 ms, the two streams are perceived as synchronized (van Wassenhove, Grant, and Poeppel 2007).[[18]](#footnote-18) (Note that there is also a difference of processing time for the different senses.)

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. As William James (1890) wrote: “The mental stream, feeling itself, must feel the time-relations of its own states.” (628) And this is one natural interpretation of what happens in the temporal illusions: in Eagleman’s example above, it is natural (though not mandatory) 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*. All that we are safe in saying is this:

1. When two events are sensed as standing in a temporal relation, it is because the perceptual system has determined (rightly or wrongly) that they do.

To summarize: When it seems to the subject as if her experiences of the said events stand in relationship *R*, we may conclude that the system has determined that they stand in this relationship. However, nothing follows about when the system first received information about these events, or even about when these events were experienced. It may often be true that the temporal structure of experience mirrors the experience of temporal structure, but there is no reason to think that this *must* be so.

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. Again: an auditory voice stream may *systematically* be experienced as simultaneous with visually experienced mouth movements when the lag is anywhere between -30 ms and +170 ms. It is unclear how “rational introspective reflection” would affect these impressions of simultaneity.[[19]](#footnote-19)

The Exportation Principle is thus not only inaccurate as stated above, but also wrong about the causation of experience. It should be amended as follows:

Exportation of Temporal Operators (Amended)

(i) 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).

(ii) They are so sensed because the perceptual system determines that they stand in this relation.

Experience of temporal order 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.” Further, our sense of time is not derived from information that impinges on the senses. Rather, it is the measure of perceived events. It is an *a priori* template for the ordering of perceived events.

Temporality is a common measure of perceived events. This is illustrated by the fact that all perceptual experience presents its objects as they are *now*, i.e., as they are at the time of the experience. This constraint operates across modalities: when you hear the song and see the dance, both are sensed as happening now, and hence simultaneously. This matching is not explained by the *facts*; it does not happen simply because the song and the dance *are* simultaneous (for even if they are simultaneous, the system’s determination “screens off” simultaneity). And itcannot be specific to vision or audition. Therefore, the representation of time cannot be modality-specific.[[20]](#footnote-20)

Obviously, all of the analogues of the *now* condition fail for experience of space. Not everything is presented as occurring *here*, i.e., where the experience is, or in any other single place. 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 perceived spatial relations of things are not projected from a modality-neutral proxy.

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: just as time is a common measure of events presented by different modalities, so 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.[[21]](#footnote-21) 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.

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 that involves bodily action such as visual scanning.

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. She 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. She 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. Much the same is true of 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.[[22]](#footnote-22)

### Isotropic Perceptual Models, Egocentric Perceptual Information

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. In the limit, 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 is a three-dimensional configuration of the parts of the jacket. The locations of things are represented relative to one another, not relative to the perceiver. They are not represented differently from different perspectives. I always perceive the jacket from a particular point of view, but my perception can be informed and interpreted in light of an isotropic model that represents how its parts spatially relate to each other.

Formally, an isotropic perceptual model consists of at least the following:

1. A set of perceivable things, *P* (e.g., parts of my jacket);
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*.[[23]](#footnote-23)

As one builds up one’s store of perceptual information concerning an object like the jacket, one’s image of it becomes progressively more complete and progressively more independent of perspective. But the information that one receives in any one view from any one point of observation will be largely egocentric. When I first began to be aware of it, the thread in my jacket manifested itself as merely an irritation of my skin. And even when I became aware of it as an external object, I still had no idea of how long it was, what colour it was, or where it was relative to the jacket. All that I knew at this point was that it was pointy and sharp, and where it was relative to the irritation of my skin. When I took the jacket off and manipulated it, I was able to place the thread relative to the shoulders and sleeves of the jacket in a way that enabled me to re-identify it even when the jacket is turned around. The isotropic model is built up from a collection of egocentrically specified information derived from one sense modality at a time. The question is how. What is the representational structure that enables active perception to construct an isotropic model?

Before we think more about this question, let us note that isotropy is ubiquitous in *retained* perceptual images. Consider iconic memory.[[24]](#footnote-24) Think of a house in which you spent many childhood years. You recall how the front hallway appears 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 that a blind person entertains.)

In retained images, 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.[[25]](#footnote-25)

A second point to consider is that my isotropic models 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. (The psychologists referred to above use the term ‘image’ to refer both to that of which we are consciously aware and the underlying mental representation. I find it more appropriate to refer to the object of conscious awareness as an image—this is always from a point of view—and the underlying mental representation as an isotropic model.)

♫♬♪

Rough face

Heavy and shiny

Isotropic model 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.

There are two transformations that must be considered with regard to the construction of isotropic models. When we experience an object, certain of its aspects are presented egocentrically. It is, for example, presented as having a back that is out of view, though the sensory characteristics of the back are out of view and unknown. Its side face might be foreshortened and it may therefore be impossible to figure out how deep it is. We may hear a sound coming from behind it, and we may not know what sort of object is emitting the sound. Isotropic models are created out of many such presentations. Their creation is a perceptualprocess of interaction, exploration, and collation. This process comes naturally to us; it is not an act of sophisticated cognition. With every movement and every shift of attention, we continually and always undertake a process that takes as some of the way to building this perspectival view into an isotropic model.

On the other hand, when we already possess an isotropic perceptual model of something, we generate egocentric views of it in the mind’s “eye” (as well as in the mind’s ear and the mind’s skin). And when we perceive a familiar thing, we experience a scenario that is pregnant with the model (or model 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. Generally, we cannot form a view of something without viewing it from a certain perspective. (This is not true of audition: I hear a band play a tune; later, I play the tune in my head, and it doesn’t seem to come from any place in the auditory field.)

Summing this up, we have:

*Isotropic Multimodality Thesis* There are retained perceptual representations are (to varying degrees) independent of the subject’s perspective or point of view. These representations are generated by “active perception,” which involves interaction with the object in order to get multiple views of it. When a visual 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. [[26]](#footnote-26)

These are 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 model—a model 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 multiple senses can act together has not, of course, escaped the attention of philosophers and psychologists. In fact, thanks to William Molyneux, an Irish writer of the late 17th century, it became one of the central problems of early modern philosophy. How are the spatial perceptions, or sensations, of different modalities related to one another? Given the cognitive non-equivalence of ideas of different modalities, how can the modalities work together? What is the nature of the information that allows information specific to one modality to be collated with that specific to 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[[27]](#footnote-27)—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.[[28]](#footnote-28)

One problem with the Privileged Modality view is that there are reasons to assign some spatial properties to one modality and some to another. As just mentioned, Berkeley thought that *distance* and *size* are represented tactually, not visually. However, when it comes to *shape*, some (including Diderot, discussed below) think that vision has an advantage.[[29]](#footnote-29) (Stokes and Biggs, this volume, take a similar view, though on different grounds.) However, distance is a component of three-dimensional shape, and so it makes no sense to assign it to a different modality than shape.

Diderot offered one canonical argument to the effect that touch does not represent shape. According to him, there isno *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. Diderot concluded that shape is a visual idea that the blind can only represent abstractly (provided that they have some knowledge of geometry). Let’s call this Diderot’s Thesis.

Now, it makes no sense to think that spatial properties can be distributed across modalities, such that some are specific to one, and others to another. 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) appeals to something like this consideration in his treatment of Molyneux’s question. He 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 when things project an image of *this* angular size when they are an arm’s length away they are large. 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 his theory. 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, on this view, 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[[30]](#footnote-30):

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 (on this view) 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—a certain “visual leftishness,” if you will. For, as Evans forcefully argues (citing 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; otherwise, there could be no coordinated multisensory awareness. And this seems to hold up empirically. Pawan Sinha’s investigation (with colleagues) of newly sighted individuals in India (Held *et al* 2011) reveals that it takes these patients a few days, 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.[[31]](#footnote-31)

Evans’s argument runs parallel to the argument about the representation of time in section II above. 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 a Common Measure argument of the same kind. 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 the scheme 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.

Common Measure arguments are, as I said earlier, extremely important. However, the Behavioural Space view suffers from some of the same difficulties as the Privileged Modality view. Suppose that, as Evans insists, the perceptual representation of space is “in terms of the behavioural dispositions and propensities to which such information gives rise” (371). 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 as coming from 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, the Behavioural Space view cannot account for the formation of isotropic images. Behavioural space is the body’s representation of space for the purposes of actions such as reaching or grasping. Such representations are egocentric. To reach for a dog’s collar, for example, I must know where it is relative to my hand at the moment of reaching. Now suppose that this was the primary representation of the dog-collar’s spatial position—i.e., where it is relative to my hand. Then, it would not be possible directly to register that the dog and its collar had retained its position relative to my hand while both had moved. For any such perception of motion would have to track both the hand and the book against a space within which behaviour occurs. Yet this is exactly how isotropic images are formed. To form such an image of a large three-dimensional object such as a house, I have to walk around it, keeping track of how I move and how the house looks as I move.

Evans was well aware of isotropic representations. He quotes Pierre Villey, “a blind Montaigne scholar who was evidently riled by the suggestion that the blind did not have genuine spatial concepts.” Villey wrote:

. . . 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. (Quoted by Evans, 369)

Yet images from any one point of view, including motor representations, are perspective dependent. If you remove the framework space in which both point of view and changes of point of view can be represented, it is hard to see how Villey’s representation could be possible. It is ironic that Evans, who emphasizes so strongly the wrongness of Diderot’s thesis, falls into exactly the same quandary.

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 increasingly 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. The virtue of this view is that it embraces a common measure of space across modality specific perceptions. 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 regardless of modality. 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 hidden 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. By analogy with the 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 models. 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 a pinprick in my finger and observe visually that my finger is in front of the TV without feeling the pinprick 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 fully 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 models 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 models—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.

Freud, S. (1899). Screen memories. *Standard Edition, 3,* 301-322. London: The Hogarth Press.

Gibson, J. J. (1962) “Observations on Active Touch,” *Psychological Review* 69: 477-491.

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.

Hoover, Adria E. N. and Harris, Laurence R. (2012) “Detecting Delay in the Visual Feedback of an Action as a Monitor of Self-Recognition,” *Experimental Brain Research* 222: 389-397.

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

Harrar, Vanessa and Harris, Laurence R. (2008) “Effect of Exposure to Asynchronous Audio, Visual, and Tactile Stimulus Combinations on the Perception of Simultaneity,” *Experimental Brain Research* 517-524.

Izmailov, Ch. A and Sokolow, E. N. (1991) “Spherical Model of Color and Brightness Discrimination,” *Psychological Science* 2: 249-259.

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.

Locke, Don (1971) *Memory* London: Macmillan.

Nigro, Georgia and Neisser, Ulric (1983) “Point of View in Personal Memories,” *Cognitive Psychology* 15: 467-482.

Mackie, John (1976) *Problems from Locke* Oxford: Clarendon 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.

Tulving, E. (1985) “Memory and consciousness,” *Canadian Journal of Psychology* 26: 1-12.

van Wassenhove, V., Grant, K. W., and Poeppel, D. (2007) “Temporal Window of Integration in Auditory-Visual Speech Perception,” *Neuropsychologia* 45: 598-607.

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

Williams, Bernard (1966/1973) “Imagination and the Self,” British Academy Annual Philosophical Lecture, 1966, published (1973) in Williams, *Problems of the Self* Cambridge: Cambridge University Press.

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, he is not exactly right. 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 relative to things in objective space changes as she moves. And they are not seen as existing outside the subject’s consciousness; they are seen as figments. 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. Fiona Macpherson (personal communication) raises the interesting question: what do two things look or sound like when they look or sound as if they are in one space? The answer must be that they appear to have space between them. Since this connecting space might not itself be seen or heard, it might be somewhat misleading to say that when the connecting space is occluded or otherwise obscured, the objects *look* or *sound* as if they inhabit a single space.The appearance of spatial connectedness would be amodal, much like the appearance of continuity of an object that is partially occluded behind an opaque picket fence—Albert Bregman (1990) discusses auditory analogues of occlusion—or the appearance of an object having a back that cannot be seen. Kant thinks that it is *impossible* to see two things as outer without them appearing amodally to be connected. [↑](#footnote-ref-3)
4. 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), who suggests that it is. [↑](#footnote-ref-4)
5. 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.” As an example, heard objects should appear to be spatially connected to seen objects. This seems right, as far as it goes. However, 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 and auditorily appear that X is to the left of Y. Secondly, there should be cross-modal “side-by-side” relations that add up to a shared Euclidean structure. These two conditions are strong enough to generate Molyneux’s problem. [↑](#footnote-ref-5)
6. Millar (2008) contains detailed synoptic discussion of empirical evidence regarding cross-modal matching. [↑](#footnote-ref-6)
7. 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-7)
8. 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-8)
9. See Izmailov and Sokolow (1991) for a recent geometrical formulation of colour similarity. [↑](#footnote-ref-9)
10. Wittgenstein: “We may call lilac a reddish-whitish-blue or brown a blackish-reddish-yellow—but we *cannot* call a white a yellowish-reddish-greenish-blue. And *that* is something that experiments with the spectrum neither confirm nor refute. It would, however, also be wrong to say “Just look at the colours in nature and you will see that it is so.” For looking does not teach us anything about the concepts of colours.” (*Remarks on Colour*, §72). [↑](#footnote-ref-10)
11. For further discussion, see Matthen 2010a. [↑](#footnote-ref-11)
12. The surface reflectance view is over-simplified—luminances and transmittances are colours too. But this does not matter for my present point. [↑](#footnote-ref-12)
13. 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-13)
14. 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-14)
15. 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-15)
16. 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-16)
17. See also Harrar and Harris (2008). [↑](#footnote-ref-17)
18. Charles Spence reports to me in correspondence that when his lab ran this experiment, a degree of confusion was observed with time lags of as much as ½ a second! (The experimenter kept halting the experiment because s/he thought that the asynchronized version had mistakenly not been played.) [↑](#footnote-ref-18)
19. 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-19)
20. Plato undertakes a similar line of thought in the *Theaetetus* (185). How do we know that a sound and a colour are different from each other, and like or unlike one another, as the case may be? Not by looking, since sounds cannot be seen. Similarly, not by listening. Plato suggests that all such cross-modal comparisons must be made by “the soul” (not the senses). [↑](#footnote-ref-20)
21. The term “active perception” as well as my characterization of the activity derive from and generalize J. J. Gibson’s (1962) notion of active touch. Earlier I used the term “sensory exploration” (Matthen forthcoming). [↑](#footnote-ref-21)
22. For further discussion of active perception, with particular reference to its epistemological force, see Matthen (forthcoming). [↑](#footnote-ref-22)
23. The pairwise distances determine the shape of a three dimensional object up to Kant’s incongruous counterparts. I shall not attempt to solve this problem here. [↑](#footnote-ref-23)
24. For more about scenes in iconic memory, see Matthen 2010b. [↑](#footnote-ref-24)
25. For the observer-field perspective distinction in memory, see Georgia Nigro and Ulric Neisser (1983). The idea was earlier noted by Freud (1899) and mentioned by Don Locke (1971, 88-89). Bernard Williams (1966/1973) makes a similar point about imagination, noticing that he is able to visualize “from the outside a figure who is myself” but also imagine things from his own point of view. [↑](#footnote-ref-25)
26. 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. Charles Spence and Tim 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-26)
27. 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-27)
28. The distance illusion is no longer accepted as an explanation of the Moon illusion. [↑](#footnote-ref-28)
29. Visual dominance is often cited in this context—the phenomenon that when vision conflicts with other modalities with regard to spatial information, the visual information is taken over by the other modalities at the expense of their own. Now, this does not show that the visual representation of space is privileged. To the contrary, it shows that each modality has an estimate of spatial properties that can be compared with the visual estimate of the same; thus, it assumes a common, cross-comparable, and transferable spatial measure. [↑](#footnote-ref-29)
30. 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-30)
31. Sinha confirms in correspondence with Alex Byrne that they did not notice any such deficiency in the newly sighted. [↑](#footnote-ref-31)