# Is Perceptual Experience Normally Multimodal?[[1]](#footnote-1)

## Mohan Matthen

The study of multimodal perception is relatively new: according to Barry Stein (2012b), the rate of publication on multisensory integration grew ten-fold in the first decade of this century. This is somewhat misleading. It was not so much that the phenomena were unknown or unstudied, but that they were not brought under the methodologies that cognitive scientists employ today to investigate multimodality. (I make no distinction in what follows between “multisensory” and “multimodal.”)

Consider speech perception. Eric Vatikiotis-Bateson and Kevin Munhall (2012) are not saying anything startling when they write:

The acoustics of speech are the signature property of spoken communication, but the movements of the individual producing the speech provide extraordinary information about the talker’s message. Head and face movements, gestures from the hands and arms, and bodily posture adjustments are part and parcel of the sequence of spoken words (421)

W. H. Sumby and Irwin Pollack found as long ago as 1954 and found (in a three-page paper published in the *Journal of the Acoustical Society of America*) that visual observation of a speaker’s facial and lip movements contribute to the intelligibility of speech, especially at low speech-to-noise ratios. The gap in the scientific literature was not so much that the visual contribution to speech communication was unknown. Rather, it was that this interaction was not connected to indicators of integrative sensory processing such as “detection and localization behaviours, speeded reaction times, enhanced/degraded speech perception, and cross-modal illusions” (Meredith 2012; see also Casey O’Callaghan’s list in section 4, this volume).

The required change of outlook came relatively slowly, for as Lynne Bernstein (2012) relates, Sumby and Pollack’s article was very little cited for a considerable period of time. (It has now been cited nearly 1800 times, but mostly in this century.) The same was true of the perceptual findings of McGurk and MacDonald (1975), now cited nearly 5000 times, and the neurological findings of Sams et al (1991) that visual perception of people talking activates the auditory cortex (this still has fewer than 400 citations.) This body of convergent evidence was raised as support for the idea that speech consists not of sounds, but of certain “gestures” of the vocal tract. But, as O’Callaghan points out in private communication, the seemingly obvious corollary that speech perception integrates auditory perception of speech sounds and of visual perception of gestures of the mouth and tongue—and that it is therefore multimodal—went almost unnoticed for quite a while.[[2]](#footnote-2)

Bernstein says that change came with Stein and Meredith’s (1993) development of theoretical models to deal with an unrelated multisensory phenomenon, the representation of space in the superior colliculus. It was in the wake of this breakthrough, that speech and other phenomena began to be treated as multisensory. Thus, Bernstein writes: “The present era of focussed research on audiovisual speech perception and its underlying neural processing arguably began in the 1990s.” She points to a special session of the Acoustical Society of America in Washington D.C. and a NATO Advanced Study Institute in France, both in 1995.

The situation in philosophy today is similar to that in cognitive psychology pre-1990. There is a host of familiar phenomena—in particular, speech perception and the perception of spatial phenomena—that philosophers simply don’t think of as multimodal. In part, this is based on a difference of focus; philosophers are concerned with perceptual *experience*, while psychologists are concerned with the interaction of sensory processes.[[3]](#footnote-3)Nevertheless, it is fair to say that philosophers’ attitudes too are founded on the absence of an appropriate theory. Many philosophers still cling to an introspectionist methodology, as well as a remnant of Empiricist Atomism that militates against the very idea of integrating distinct sources of information in perception.

My main aim in this paper is to show how sensory integration is the norm in perception, even within single modalities such as vision. I contend, in short, that the alleged problems of *cross*-modal integration, which arise for example in discussions of Molyneux’s Question, actually have nothing to do with the participation of multiple modalities. But—and this is worth saying explicitly—my concern here will be with experience, not sensory systems, structures, or processes. I am sceptical about how much one can know on the basis of introspection. Nonetheless, it is experience that I will be talking about—my claims about the content of experience find support in its effect on action-capacities, and does not rely on personal introspection, as I will document in the appropriate places.

The atomistic foundations of philosophical resistance to sensory integration must be re-evaluated. When they are, we will come to see perception as “richly multisensory.” The term is O’Callaghan’s, and I’ll offer a definition distilled from his writing:

A multisensory experience is *richly multisensory* if (a) it is not co-consciousness of separate unisensory states, and (b) its content is not the mere conjunction of the content of unisensory states.[[4]](#footnote-4)

I will concentrate on showing that our experience of spatial relations is richly multisensory. O’Callaghan, this volume, considers other perceptibles, such as flavour.

### Place

1. Think about these examples:

(a) There is a stale smell in the laundry room. You sniff, moving about the space. Ah, you’ve found it. It’s the wet dishtowel that fell into the crack between the washer and dryer. (cp. Von Békésy 1964)

(b) You are listening to an orchestra. A single woodwind enters. Who is it? Scanning the orchestra, you see the clarinettist fingering her instrument and moving in time. Now you hear the music as coming from her place. (cp. Allais and Burr 2004)

In both scenarios, an imprecisely located sense-feature acquires precise location by the participation of another modality—the sense of self-motion (which in turn employs the vestibular and proprioceptive systems and motor feedback) as well as vision in (a), and vision as well as proprioception (to scan the orchestra) in (b). The modalities work together by first detecting causation; the clarinettist is (perceptually, not cognitively) identified as the source of the sound because her visual movements coordinate with the sound; the dishrag is identified as the source of the odour because smell intensity-gradients radiate from it. Once an object evident to one modality is identified as the cause of an event evident to another, the spatial location of the two modalities is coordinated. Once the visible movements of the clarinettist are identified as the source of the clarinet’s audible melody, the latter is localized in coordination with the former. So we have an earlier (supposedly) unimodal experience and a later transformed experience that contains source and spatial information provided by several modalities working together. The transformed experience is richly multisensory: the earlier unisensory experience did not contain source or detailed spatial information. The earlier experience has disappeared and is not a co-conscious component of the transformed experience.

In normal perception, we are presented with a perceptual image that consists of material objects, sounds, smells, and various tactile qualities. I hear the banging of pots in that brightly lit restaurant kitchen from which all those delicious smells and radiant heat emanate. This is the kind of image that perceptual experience delivers. In it, all of the perceptual features are bound together and connected by spatial and causal relations. Perhaps some (though clearly not all) of these features arise from single modalities. However that might be, their causal and spatial relations are delivered multimodally. The spatially unified and causally connected image of our surroundings arises from richly multisensory experience.

Vision has special responsibility for spatial representation because it provides the finest grained representations; in (b) visual spatial representations localize sounds with a precision that surpasses audition. O’Callaghan writes that in cases like these, cross-modal processes “improve the accuracy and reliability of the overall perceptual result.” (Cp. O’Callaghan 2014, 148, esp. n 15.) I agree, but I want to emphasize an additional point—that these processes enhance the fineness of spatial grain of the perceptual image *regardless of accuracy and reliability*. As well, there are the causal connections. As O’Callaghan observes, there is a difference between noticing that one and the same object is red and rough and noticing that a visually identified object is red and that a tactile object is rough. There is also a difference between seeing a woman and hearing a clarinet, and hearing a woman you see playing the clarinet. The latter is a richly multisensory experience that is additionally *constitutively* multimodal, not merely causally so. (See Connolly 2014 for discussion of constitutive multimodality.)

The perceptual image is a multimodal melding of information derived from the senses. It is an organic spatial whole, not a mere conjunction of unimodal features.

1. One point should be emphasized at the outset: the perceptual representation of space is not specifically visual or specifically bodily. It is a template or matrix into which all the modalities place features. The same object is perceived as possessing many features from several perceptual modalities. To deliver such a presentation, vision and touch must have access to the same spatial matrix.[[5]](#footnote-5) It follows, first, that as I have argued elsewhere (Matthen 2014):

The representation of space is *pre-modal*; it is, in Kant’s words, an *a priori* intuition.

In light of this, I want to urge, second, that:

The total perceptual image is multimodally coordinated.

(For the sake of clarity, it is the second proposition that is more important in the present context, though the first provides important background.)

The representation of space underlying conscious perception of the external world is objective and allocentric. Perceptual consciousness *seems* like a photograph or a movie: a view from a certain perspective with no objective indication of *where* in the world the point of view is from. Am I looking up Church Street or Spadina Avenue? Am I looking north or west? The intuition based on phenomenology is that perception cannot answer such questions; perceptual phenomenology is egocentric, or so it seems. This intuition has been shown to be wrong: we now know that the perceptual image is linked to something more like the map display on your GPS—an objective map that has your position and heading marked. When you are in a familiar setting, you are implicitly capable of relating the objects you perceive to landmarks beyond the perceptual horizon (Epstein and Vass 2014). Your experience of the room in which you now stand is, for example, pregnant with the location of the unseen corridor outside.[[6]](#footnote-6)

I said earlier that the perceptual representation of the spatial matrix is pre-modal. The brain uses this matrix, which happens to be located in the hippocampal formation, to make a map of the locales that you inhabit by correlating movement with perceptual features. The matrix is non-empirical and a priori, and it is filled by the collaboration between the sense of self-movement and other perceptual modalities. (*Historical note.* E. C. Tolman 1948 was the pioneer of cognitive maps. O’Keefe and Nadel 1978 is the now canonical scientific treatment, with interesting allusions to Kant. The Karolinska Institute’s announcement of the Nobel Prize for Physiology, 2014, awarded to John O’Keefe and Edvard and May-Britt Moser was explicit: the hippocampal coding is “a comprehensive positioning system, an inner GPS.” It also invokes the Kantian echo. It is instructive to think about the perceptual infrastructure needed to maintain this kind of cognitive structure.)

### Philosophical Resistance

1. Many philosophers are sceptical about multimodal experience. Fiona Macpherson (2011) describes their attitude in the following way:

A common assumption was that the sensory modalities were perceptual systems isolated from each other. Each sensory system produced, unaffected by the others, a perceptual experience characteristic of that sensory modality (a ‘uni-modal’ experience) and perhaps other uni-sensory, non-conscious, sub-personal representational informational states characteristic of the modality.

What is the basis of this “common assumption”?

One kind of approach, following H. P. Grice (1962), is based on the idea that each modality produces experiences of a characteristic phenomenal character. For example, although hearing somebody say ‘bat’ is phenomenally different from hearing somebody splash about in the ocean, the two experiences share a generic auditory phenomenology that marks both off from *seeing* somebody say ‘bat’ or *seeing* somebody splash about in the ocean. Analogously, the claim about (a) and (b) is that the transformed experiences are phenomenally like the earlier ones in the crucial modality-identifying way. The stale smell *feels* smelly in both the first and the transformed experience of (a); both (b) experiences *feel* auditory.[[7]](#footnote-7)

It is fair to say that this approach runs against the scientific current. In perceptual science, the modalities are thought of as *sources* or *processes*, not *products*. Perception is a process that results in an image of the organism’s surroundings. In order to construct such an image, animals sample energy. Since there are many kinds of energy—electromagnetic, acoustic, thermal, mechanical, chemical—many different sampling mechanisms have evolved. Generally speaking, there are different *transducers* for different kinds of energy—the cells that convert light into neural signals are different in kind from those that do the same for sound, chemical reactions, etc. The processes that extract ecological content from these neural signals are also specialized—at least in the early stages of processing, which are energy specific. These specialized and separate processes are the cognitive scientist’s modalities. Later in the perceptual process, representing the environment begins to be the focus (rather than analysing patterns in the specialized receptors), the different modalities contribute to a common conversation. This gives rise to the simplified schema of Figure 1, which assumes that the data-streams stay separate in the early stages of processing, and that there is no multimodal influence early on. The current recognition of multisensory processing rests on the dawning realization that these data-streams inter-communicate from the very first stages.

*AUDITORY TRANSDUCERS (COCHLEA)*

VISUAL DATA PROCESSING

SOUND PATTERN INFORMATION

LIGHT PATTERN INFORMATION

CONTENT ABOUT OBJECTS, SOUNDS, AND SOUND SOURCES (etc.)

OTHER MODALITIES

*LIGHT*

*SOUND*

*AUDITORY DATA PROCESSING*

*MULTIMODAL PERCEPTUAL PROCESSING*

*VISUAL TRANSDUCERS (RODS AND CONES)*

Figure 1: The Modalities as Sources for Multimodal Perceptual Processing. It is worth noting that in fact visual information about speech enters the auditory stream very early. (Sams et al 1991)

Of course, the idea of a “source” is context dependent, and the scientific concept of modality is correspondingly flexible. Touch, for example, is often said to be multisensory, on the grounds that it has transducers for different kinds of energy—stretch receptors, pain receptors, thermal receptors. But it can also be treated as a single modality on the grounds that it integrates the information coming from these diverse receptors at a very early stage. (Fulkerson 2013 has an excellent discussion.) Flavour perception, mainly driven by receptors in the tongue and the nose, is the opposite sort of case. Both sets of receptors respond to chemical signals, but since the nasal receptors operate independently of those in the tongue, it is common to distinguish (retronasal) olfaction from taste and to say that these come together in flavour perception, which is usually held to be multisensory (Smith 2015).

In perceptual science, the question posed in my title is about the depth and relative isolation of the supposedly separate processes in Figure 1. Are the streams well integrated right from the start, or are they integrated only much later?

1. As I have described the dialectic, philosophers distinguish the modalities by differences of phenomenal character, while perception scientists do so by differences in the way that informative content is extracted from different energy sources. Who is right? Does it even make sense to ask this question?

One approach to the question is *criterial*. The idea is that any consistent criterion marks off a legitimate category. Here’s an extreme example. One criterion associated with the word ‘bank’ makes it true of financial institutions; another comprehends the shores of rivers. These are different, but neither is wrong. Similarly, some say, with modality. Philosophers use phenomenal character to demarcate the senses; they operate with a phenomenological concept. Scientists use process; they operate with a different concept. These are different, but both are legitimate; because each is consistent, each defines a possibly non-empty class. This is the basis of the criterial approach.

Two philosophers who align themselves with the criterial approach, though in different ways, are Louise Richardson (2013) and Fiona Macpherson (2011, 2014). Neither adopts quite as extreme a position as the one outlined above—both think there is *some* overlap of meaning among different criteria—but both hold that there is no conflict among various criteria that define the modalities. Richardson holds that there is an “everyday” conception of flavour on which it is a single modality. Since this concept is defined by different criteria, it is not challenged by the scientific concept, which (as noted above) treats flavour perception as multisensory. These are simply different ways of carving the roast, she suggests. Macpherson (2011) also thinks that the phenomenological and process criteria are different and orthogonal—she reserves the term “multimodal” for the first and “multisensory” for the second. (As a reminder: I use these terms interchangeably.) She follows Grice in delineating four ways of assigning a given perceptual state to a modality, and insists that “we should not choose between these criteria for individuating the senses but instead use them all.” Macpherson (2014) combines the pre-existing criteria into a single similarity space for modalities.

A second approach to defining the modalities is *analytic*. The assumption here is that modalities are real psychological processes, and that the task is to discover what differentiates them. (Matthew Nudds 2004 explicitly *denies* that the distinction between modalities is based on “nature.”) On such an approach, one might begin the consideration of cases (a) and (b) with the idea that vision, audition, olfaction, and the sense of bodily self-awareness do different jobs, and try, as a first step, to figure out how these jobs differ. Then, one might try to formulate a general partitioning of perception in ways that capture relevant differences. The schema of Figure 1 is an example of what might result from this procedure.

On this reckoning, the phenomenological and the process-oriented descriptions are different ways of partitioning perceptual processing into its natural parts. They might turn out to be incompatible, or complementary, or incoherent—this depends on other features of the general scheme. This was Grice’s approach to the question (and Aristotle’s, for that matter). Analytic inquiries are open-ended. They *could have* revealed that there are no modalities in anything like Richardson’s everyday sense, or that the distinction depends on explanatory context, as Nudds suggests. These would be significant insights. In the event, though, they came up with something like Figure 1. (See, for example, Stein and Meredith 1993, Introduction.)

In my view, the criterial approach makes too many questionable assumptions about the domain of investigation. For example, Macpherson (2014) proposes that the criteria are independently conceived axes for a combined similarity metric, from which Principal Component Analysis (PCA) can derive a taxonomy. There is no reason to accept this. Suppose you charted foods by sweetness, softness, colour, “height” of cuisine, distance between towns of origin, and price per ounce. Would PCA yield up a useful categorization of food? Macpherson hopes that PCA will reveal “an extra deep layer of information” “hidden” within the similarity space of the modalities. This overestimates the efficacy of statistical methods. PCA is designed to pick out statistically independent axes of variation. Sometimes such axes correspond to, or reveal, significant underlying measures; equally often, PCA yields meaningless results. Properly understood, PCA serves as a preliminary to analytic investigation. It is not a panacea, and its mere existence is not a good reason to accept even the partial legitimacy of multiple criteria. Nor, *pace* Richardson, is the mere existence of an everyday conception of flavour any reason to believe in its coherence or utility.

As must be obvious, I favour an analytic approach. I can’t argue for this here, but let me simply observe that it is hard to see how the phenomenological criterion can yield anything useful by way of determining whether the transformed experiences in (a) and (b) are multimodal. This criterion directs us simply to introspect these experiences and decide whether they “feel” purely auditory or purely olfactory or mixed. Such a procedure is far too poorly specified to be useful in contested cases. Does the transformed experience of the clarinet *feel* auditory? Perhaps, but would we say it also felt visual when we come to notice that it depends on keeping our eyes open? It is at least possible that we would.

We need a far more nuanced approach.

### Integration

1. Locke’s negative answer to Molyneux’s question is motivated by empiricist atomism: the newly sighted man “has not yet attained the experience, that what affects his touch so and so, must affect his sight so and so.” For Locke, the ‘so and so’ stands for what we would call phenomenal character. To say that a sphere is known by touch is for him to say that we know that *a sphere* causes the phenomenal character produced by touching a sphere. (Locke’s position is implicitly accepted by contemporary philosophers, like Grice, who say that the modalities are defined by phenomenal character.)

In quasi-Fregean terms (“mode of presentation”), Locke’s reasoning can be represented as follows:[[8]](#footnote-8)

*Atomism* A simple perceptual experience is, in virtue of its phenomenal character, a mode of presentation of a single perceptual feature.

*No Synonymy* Perceptual experiences that have different phenomenal character are different modes of presentation.

*Cross-Modal Difference* Any two perceptual experiences from different modalities have different phenomenal characters.

*Conclusion 1 (Unimodality)* Since multimodal perceptual experience must incorporate multiple phenomenal characters (by Cross-Modal Difference), it must be composite.

*Conclusion 2 (Negative Response to Molyneux)* Since experiences of different modalities are different modes of presentation (by No Synonymy), it is contingent whether they present the same or different features.

*Conclusion 3 (Decomposing the Multimodal)* Since visual and tactile sphericity are different (by Atomism and No Synonymy), looking at a sphere you hold in your hand, creates a conjunction of separate experiences: you are aware of visible sphericity; you are aware of tactile sphericity; you are co-conscious of both. Similarly, when you notice that it is red and rough.

Locke’s argument makes it pointless to pursue questions about multimodality: it is a definitional stop. This is one reason we urgently need to reassess it.

Now, it is unclear how an atomist should interpret the experience of space. As Kant realized, every *perceptual* experience (i.e. every “outer” perception) is spatial: for example, every perception of *red* is experience of *red in a place*. How should this be understood? We cannot say that experience of feature-in-place is composite—it isn’t experience of feature *plus* experience of place. This separatist account would posit awareness of red-not-in-place as prior to awareness of red-in-place. Perceptual experience of red-not-in-a-place seems impossible, especially as a norm. (Possibly you could have a fleeting non-perceptual experience of red-not-in-a-place, for instance when you press down on your eyeball.) Moreover, the account does not tell us how to tie feature and place together: if perception of place is separate, it cannot be the “binding parameter.” Atomism is on shaky ground. It cannot account for the integration of features in a spatial field.

Examples (a) and (b) suggest that the modalities cooperate to localize features, possibly with vision dominating (Millar 2008, Stokes and Biggs 2014). Such cooperation would not be possible if there were not a shared “vocabulary” of space among the modalities. This shows that even simplest experiences have a locational component they share with experiences of other modalities. Nor is this sharing learned. Even newborns turn toward a sound and track a moving object with their eyes.[[9]](#footnote-9) Moreover, they turn their heads away from *odour* sources.[[10]](#footnote-10) No Synonymy is questionable with respect to our experience of location.

1. The same goes for direction. Gareth Evans (1980) wrote: “Disembodied perceivers would have no notion of up, down, left, right, forwards, and backwards.” It is important to temper this observation with the realization that perspective is integrated with allocentric representation. If you stand on your head, lower parts of the environment project to the chin-ward end of the retina; nevertheless, you continue to experience the up-direction as away from the Earth. *Up* is always defined by awareness of gravity, even visually. (Imogen Dickie tells me that rock climbers always understand ‘up’—as in ‘Look up,’ ‘Reach up’—as away from the Earth.) Our sense of *up* is bodily, even in visual presentations. A newly sighted man should be able to tell which way *looks* up.

Even more importantly, the hippocampal “GPS” incorporates direction cells. These indicate the perceiver’s orientation in a familiar scene; there are cells, for example, that will fire when the perceiver is pointed north, and others that fire when she is headed south. These cells subserve a sense of direction informed by past experience in all modalities. As O’Callaghan (this volume) writes: “an experience that is associated with one modality on a given occasion can depend upon past or present sensory episodes that are associated with another modality.”

1. The phenomena we have been discussing illustrate a more general principle of perception that applies both within and across modalities:

*Integration* Experience of any given sense feature is modulated by extracted content regarding other sense features. Thus, experience of any one feature contains information about other features.

For a within-modality example, take the experience of colour. You see coloured shapes: the colours are tightly confined within the contours. For instance:

Now, as Margaret Livingstone (2002) writes, “cells that code color have larger receptive fields, and there are fewer of them than the cells used for resolving contours” (95). The visual system renders the spatial location and extent of the *red* that you see above far less precisely than the triangular contour (for which it uses brightness, a component, but not the whole of colour). How then does the colour cling so exactly to the form?

The visual system makes colour boundaries in two steps. The first step is described in Anne Treisman’s feature integration theory (Treisman and Gelade 1980). Treisman realized that colour and shape are processed separately in the brain; these processes proceed in parallel, one resulting in a spatial map of colours, the other of forms. In order to co-locate shapes and colours, the two maps have to be collated. Treisman and her co-workers were able to demonstrate that this requires a third process, which she identified as *attention*. In less attended parts of a scene, perceivers cannot reliably co-locate separately processed features. Attentive co-location is followed by a second step that Livingstone articulates: the system “codes the color of an object by signalling [the object’s] edges, . . . letting the color ‘fill in’ perceptually” (99). In this respect, colour-form perception is like the audio-visual location of the clarinet’s melody. In both cases, a finer determination of place replaces a rough one.

Perception performs these integrative operations because it is adapted to certain environmental regularities. It confines colour within object contours because it is extremely rare to find colour that spreads from one object to another. Thus, visual experience is not simply of exactly coterminous colour and shape, but that of a coloured shape. Experience of a coloured surface is different from that of a colour and a surface; the integrated experience is different from its separate atoms. (This contradicts the theory that David Lewis 1966 attributes to “British empiricism and introspectionist psychology”: “the content of visual experience is a sensuously given mosaic of color spots, together with a mass of interpretive judgements injected by the subject.”) Atomism asserts that one feature carries no information about another. But here we find that shape carries information about the extent of colour.

### Conclusions

Cognitive scientists now accept the existence of multimodal perception, but philosophers are still reluctant to do so. On occasion, this difference rests on the use of different criteria. Take touch. As mentioned earlier, most psychologists are inclined to say that it is multimodal, on the grounds that it involves a variety of receptor types sensitive to different types of incident energy. Philosophers are more reluctant to allow this, because they use other criteria. For example, Matthew Fulkerson (2013) has argued that touch is a single modality because it co-assigns its features to the same objects. This is an example of a fruitful controversy about criteria.

In this paper, I have probed a less productive argument that some philosophers use to cast doubt on multimodality. It is the idea that the features presented to us in experience are logically independent of one another (Atomism) combined with the stipulation that any single non-composite experience has indivisible phenomenal character. The conclusions drawn from these two premises are first that the phenomenal character of a non-composite experience can only belong to one modality, and second, that the feature presented by such an experience cannot depend on any other feature.

My challenge to this line of thought has been two-fold. First, I have claimed that it cannot accommodate the essentially spatial character of external position, because the representation of space takes input from several modalities. Second, I have argued that Atomism is misguided. The representation of sensory features generally takes content from the representation of other sensory features, particularly with regard to space (and also time, though I did not discuss this here)—and this happens across modalities just as often as it does within modalities.

In the end, some philosophers may stick with Atomism, preferring to reformulate the kinds of integration that I have described in terms that make them compatible with Atomism. For a philosopher who chooses this option, two questions remain. The first concerns feature-in-place: how can this be made compatible with Atomism? Second: How will the criterion of multimodality be adjusted so as not to rule it out as incoherent right from the start? One way or another, it would be healthy to set the issue up in such a way that permits a conversation with cognitive scientists.

I have not discussed some interesting phenomena that come up in philosophical discussions of multimodality, in particular speech and flavour. I have stuck with space. This is both because multisensory integration is easier to establish in the spatial case, and because I believe that the principles of multisensory integration in spatial perception will find a natural extension to these other domains. In psychology, as we noted in the beginning, investigation of spatial maps was a gateway to the discussion of other multisensory phenomena. Perhaps the same can happen in philosophy.[[11]](#footnote-11)

## REFERENCES

Allais, David and Burr, David (2004) “The Ventriloquist Effect Results from Near-Optimal Bimodal Integration,” *Current Biology* 14: 257-262. DOI 10.1016/j.cub.2004.01.029

von Békésy, Georg (1964) “Olfactory Analogue to Directional Hearing,” *Journal of Applied Physiology* 19: 369-373.

Bernstein, Lynne E. (2012) “A Historical Perspective,” in Stein 2012a: 397-405

Connolly, Kevin (2014) “Making Sense of Multiple Senses,” in R. Brown (ed.) *Consciousness Inside and Out: Phenomenology, Neuroscience, and the Nature of Experience* Springer

Epstein, Russell A., and Vass, Lindsay K (2014) “Neural Systems for Landmark-Based Wayfinding in Humans,” *Philosophical Transactions of the Royal Society B* 369: 20120533. DOI: 10.1098/rstb.2012.0533

Fulkerson, Matthew (2013) *The First Sense* Cambridge MA: MIT Press.

Lewis, David K. (1966) “Percepts and Color Mosaics in Visual Experience,” *Philosophical Review* 75 (3): 357-368.

Livingstone, Margaret (2002) *Vision and Art: The Biology of Seeing* New York: Abrams

Macpherson, Fiona (2011) “Cross-Modal Experiences,” *Proceedings of the Aristotelian Society* CXI: 429-468.

Macpherson, Fiona (2014) “The Space of Sensory Modalities,” in Stokes, Matthen, and Biggs 2014: 432-461.

McGurk, Harry and MacDonald, John (1976) “Hearing Lips and Seeing Voices,” *Nature* 264: 746-748

Matthen, Mohan (2005) *Seeing, Doing, and Knowing: A Philosophical Theory of Sense Perception* Oxford: Clarendon Press.

Matthen, Mohan (2014) “Active Perception and the Representation of Space,” in Stokes, Matthen, and Biggs 2014: 44-72.

Matthen, Mohan (2015) “Individuating the Senses,” in M. Matthen (ed.) *The Oxford Handbook of the Philosophy of Perception* Oxford: Oxford University Press.

Meredith, M. Alex (2012) “Multisensory Convergence: Where It All Begins,” in Stein 2012a: 3-11.

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

Nudds, Matthew (2004) “The Significance of the Senses,” *Proceedings of the Aristotelian Society* CIV: 31-51.

O’Callaghan, Casey (2014) “Not All Perceptual Experience is Modality Specific,” in Stokes, Matthen, and Biggs: 133-165.

O’Keefe, John and Nadel, Lynn (1978) *The Hippocampus as Cognitive Map* Oxford: Oxford University Press.

Richardson, Louise (2013) “Flavour, Taste, and Smell,” *Mind and Language* 28 (3): 322-341.

Richardson, Louise (2014) “Space, Time, and Molyneux’s Question,” *Ratio* (new series) XXVII (4): 483-505.

Rieser, John, Yonas, Albert, and Wikner, Karin (1976) “Radial Localization of Odors by Human Newborns,” *Child Development* 47: 856-859.

Sams, M., Aulanko, R., Hämäläinen, M., Hari, R., Lounasmaa, O. V., Lu, S. T., & Simola, J. (1991). “Seeing Speech: Visual Information from Lip Movements Modifies Activity in the Human Auditory Cortex,” *Neuroscience Letters* 127 (1): 141-145.

Siewert, Charles (1998) *The Significance of Consciousness* Princeton: Princeton University Press.

Smith, Barry (2015) “The Chemical Senses,” in M. Matthen (ed.) *The Oxford Handbook of the Philosophy of Perception* Oxford: Oxford University Press.

Stein, Barry E. (ed.) (2012a) *The New Handbook of Multisensory Processing* Cambridge MA: MIT Press.

Stein, Barry E. (2012b) “Introduction,” in Stein, Barry E. (2012a): xi-xvi.

Stein, B. E. and Meredith, M. A. (1993) *The Merging of the Senses* Cambridge MA: MIT Press.

Stokes, Dustin and Biggs, Stephen (2014) “The Dominance of the Visual,” in Stokes, Matthen, and Biggs 2014: 350-378.

Stokes, Dustin, Matthen, Mohan, and Biggs, Stephen (eds) (2014) *Perception and Its Modalities* New York: Oxford University Press

Sumby, W. H. and Pollack I (1954) “Visual Contributions to Speech Intelligibility in Noise,” *Journal of the Acoustical Society of America* 26 (2): 212-215.

Tolman, E. C. (1948) “Cognitive Maps in Rats and Men,” *Psychological Review* 55 (4): 189-208.

Treisman, Anne M. and Gelade, Garry (1980) “A Feature-Integration Theory of Attention,” *Cognitive Psychology* 12: 97-136.

Vatikiotis-Bateson, Eric and Munhall, Kevin G. (2012) “Time Varying Coordination in Multisensory Speech Processing,” in 2012a: 421-433.

1. My title plays on Casey O’Callaghan (2014): “Not All Perceptual Experience is Modality Specific.” [↑](#footnote-ref-1)
2. Initially, I found O’Callaghan’s observation incredible: how is it possible *not* to remark on multisensory perception in the context of the Motor Theory of Speech Perception? Then I found that *I* had cited Sumby and Pollack, and McGurk and Macdonald without a single mention of multimodal perception (Matthen 2005, 220). Thanks, Casey! [↑](#footnote-ref-2)
3. Stein and Meredith (1993, 1) say: “Of primary concern are observations that the use of one sensory system can influence (e.g., enhance, degrade, change) perception via another sensory system.” [↑](#footnote-ref-3)
4. Cp. O’Callaghan, this volume, section 2, end: “a multisensory perceptual experience may have features beyond those that are associated on that occasion with each of the respective modalities and that do not accrue thanks to simple coconsciousness.” [↑](#footnote-ref-4)
5. Touch and proprioception also operate with an egocentric “body schema.” This is, among other things, an independent spatial framework that has to be cross-calibrated with the allocentric spatial matrix that vision contributes to. [↑](#footnote-ref-5)
6. Cp. Charles Siewert (1998, 258): “To recognize one’s surroundings is to see them as the gateway to somewhere else . . . The temporarily disoriented person may see where things are around him accurately, but not recognize his surroundings—and when he does, the character of his experience changes.” [↑](#footnote-ref-6)
7. The account I favour is related, but different in a crucial place: an experience is visual if it results from characteristic activities involving the eyes; it is auditory if we use the ears in characteristic ways, etc. (Matthen 2015). This has something in common with the phenomenological account; I can feel that I am using my eyes, so the resulting experiences are felt to be visual. But unlike that account, particularly Grice’s version, mine does not rule out the possibility that visual perception might also be auditory. [↑](#footnote-ref-7)
8. Richardson (2014) offers a somewhat similar reconstruction of Locke’s thinking. [↑](#footnote-ref-8)
9. Richardson (2014) argues that “whilst the spatial structure of perceptual experience differs across the senses, its temporal structure does not.” She contrasts shape with rhythm: it is evident to the senses that seen rhythm is the same as felt rhythm, but it is not evident to the senses that seen shape is the same as felt shape. What would she say about *place*? Isn’t it immediately evident that the place a sound comes from is the same as the place you see a synchronous flash of light? The evidence of infants and of the newly sighted suggests that cross-modal identification of place *is* immediate. [↑](#footnote-ref-9)
10. Rieser, Yonas, and Wikner 1976: “The odor source was placed near the nose slightly to the left or right of midline, with its position randomized over repeated trails. . . It was found that as a group, the newborns turned away from the odor source more frequently than they turned toward it” (856). I’d like to say this implies a positive answer to MolyNOSE problem. [↑](#footnote-ref-10)
11. Warm thanks to Casey O’Callaghan and Kevin Connolly, both of whom helped me understand the basic issues. [↑](#footnote-ref-11)