This is a pre-copyedited, author-produced PDF of an article accepted for publication in *British Journal for the Philosophy of Science* following peer review. The version of record Alisa Mandrigin, Multisensory Integration and Sense Modalism, *The British Journal for the Philosophy of Science*, 72:1, pp. 27-49 is available online at: https://doi.org/10.1093/bjps/axy070

Multisensory Integration and Sense Modalism

Alisa Mandrigin

Abstract

The Bayesian model of multisensory cue integration proposed by Ernst and Banks [2002] provides an attractive model for understanding a way that our sensory systems may interact. Moreover, it has been suggested that the process of multisensory integration that it models underpins conscious experiences with multisensory representational contents merged across modalities (de Vignemont [2014b]). Should we therefore take empirical support for the Bayesian model as evidence of the multimodality of perception? Focusing on evidence of integration across vision and touch, I argue that apparent support for the model does not warrant the rejection of the view that each of our conscious perceptual experiences is associated with one and only one sense modality.

1 Sense Modalism and the Independent Systems View

The sense modalities are ways of perceiving. We typically think of ourselves as having five distinct ways of perceiving: we see, hear, touch, taste, and smell. And we typically think of our conscious perceptual experiences as being modality-specific. That is, our conscious perceptual experiences will be visual, or auditory, or tactile, and so forth. I will call the view that each of our conscious perceptual experiences is associated with or assignable to one and only one sense modality Sense Modalism.

Sense Modalism is not the view that at any time we have at most a single conscious perceptual experience belonging to one sense modality. As I'm typing on my keyboard I feel the keys beneath my fingertips, I hear the tapping sound produced, and I see the characters appear before me on the computer screen. Moreover, it's not simply the case that I seem to be able to perceive a number of distinct entities, each through a distinct sense modality, at the same time. I can look at my coffee cup at the same time as holding it in my hand, thus seeing and touching the same object at the same time. Casey O'Callaghan ([2014a], [2014b], [2017]) has labelled the view that our perceptual experiences are phenomenally unified or co-conscious the 'minimal multisensory view'.[1] The suggestion that modality-specific experiences are co-consciousness in this way is not problematic for the Sense Modalist, for she can assert that we can have a number of distinct conscious perceptual experiences, each of which belongs to a different sense modality, at the same time. We might not have experiences associated with each sense modality at all times, but at a time we often have a number of concurrent or temporally overlapping experiences, each one associated with a different sense modality. I see the cup, smell the coffee brewing, feel the carpet beneath my feet and hear the birds singing outside, all at the same time.

One way in which we might call into question the traditional conception of perceptual experience with which we began is by arguing that we have more than five sense modalities. So, for example, we might argue that we have two ways of perceiving through touch: a passive sense of touch and an active sense of touch (Bayne [2014], p. 16). Or, we might argue that our ordinary conception of taste is mistaken and that, in addition to the sense of taste proper, we have a distinct way of perceiving flavour, which combines taste with retronsal olfaction and touch in the mouth (Auvray and Spence [2008]; Smith [2015]).

But there is another way in which we can question the traditional conception of perceptual experience: by rejecting Sense Modalism. A number of philosophers have recently argued that conscious perceptual experience is or can be 'multimodal' (de Vignemont [2014a], [2014b]; Nudds [2014]; Bayne [2014]; O'Callaghan [2014a], [2014b], [2017]).[2] One initial reason for finding this second kind of claim appealing is that we have a large body of evidence indicating that the sensory systems interact with one another.

Perceptual experience is the result of the processing of sensory information in the brain arising from the stimulation of sensory receptors. Typically cognitive scientists have adopted the strategy of investigating perception by studying one sensory system at a time in isolation from the others, where sensory systems are individuated by receptor type and medium of detection. This approach assumes what I will call the Independent Systems View: the view that the sensory systems operate entirely independently from one another.

There is now a large body of evidence that shows that the Independent Systems View is mistaken. Here is just a sample. Cross-modal correspondences are pairs of apparently unrelated stimuli detected by different sensory systems that we reliably match together when prompted and that can influence our responses to tasks. For example, when asked we reliably match large objects with low pitch sounds and small objects with high pitch sounds. Gallace and Spence ([2006]) used a speeded visual size discrimination task to test whether the cross-modal

---

[1] Spence and Bayne ([2013]) have argued that, to date there is no decisive empirical evidence to support the claim that we perceive through more than one sense modality at a time. They argue that the successful performance of any task that might appear to show that we can, for example, see and hear at the same time, may be explained instead in terms of 'quick switching' between visual and auditory experiences. The question of how we might endeavour to show that we can have a number of conscious perceptual experiences at the same time is an interesting one, but not one that I will pursue further here.

[2] In at least some cases the question of whether we have more than five sense modalities is not independent from the question of whether conscious experience is multimodal. For example, there is an ongoing debate about whether we should think that we have an additional sense modality for the perception of flavor or, rather, that we have multimodal experiences of flavour properties (Auvray and Spence [2008]; Smith [2015]; Briscoe [2016]; O'Callaghan [2017]).

correspondence between pitch and size would impact response times. The task was to indicate whether the second of two discs, presented one after the other, was smaller or larger than the first. They found that participants responded more quickly when the second disc was presented synchronously with a 'congruent' sound: a small disc with a high-pitched sound, or a large disc with a low-pitched sound.

Neurophysiological studies in animals have found multisensory response enhancement and supression effects: that is, the signals generated by some cells in what have come to be recognised as multisensory areas of the brain in response to multisensory stimulation are greater or smaller than the individual component responses (see (Stein and Meredith [1993]) for a review).

Cross-modal illusions also show that the sensory systems interact with one another. For example, when subjects are presented with spatially discrepant visual and auditory stimulation at the same time, participants mislocate the auditory stimulus in the direction of the visual stimulus. This is the spatial ventriloquism effect (Bertelson [1999]). The sound-induced flash illusion is another example of a cross-modal illusion. When subjects are presented with two beeps at the same time as a single light flash, around sixty per cent of participants report not only hearing the two beeps, but also seeing two light flashes instead of one (Shams *et al*. [2000]). In the parchment skin illusion, amplification of the sound produced when we rub our hands together results in us experiencing our own hands to be drier (Jousmaki and Hari [1998]); in the McGurk Effect, subjects are presented with a video of lip movements that produce the phoneme /ga/ with the phoneme /ba/ dubbed on to it, resulting in the perception of the phoneme /da/ (McGurk and MacDonald [1976]).

So much, then, for the Independent Systems View (Briscoe [2016], p. 1).[3] Is the evidence that indicates that much sensory processing is multisensory sufficient to show that the Sense Modalist is mistaken? It might be tempting to assume that the answer to this question must be in the positive. However, Sense Modalism is a claim about conscious perceptual experience. In order to reject Sense Modalism on the grounds that experience is multimodal, we need an argument for why we should think that at least some perceptual experiences are assignable to more than one sense modality.

2 Multimodal Perceptual Experiences

There are a number of ways to go about arguing for the multimodality of conscious perceptual experience. One line of argument has focused on the fact that some of the properties and the objects we perceive can be perceived through more than one sense modality. So, for example, we come to be aware of location through vision, audition and touch. And we can perceive material objects through both vision and touch. The claim is that, at least some of the time, our experiences of such properties and objects are not modality-specific, but multimodal. My conscious experience of the height of the cup is not visual or tactile, but visuo-tactile. What does this mean?

I'm going to assume that perceptual states have representational content. Let's say that I see a cup that is in front of me. At the present time the only sensory information I have about the cup is visual. I have a visual experience of the cup, and that experience has representational

[3] See (Calvert *et al*. [2004]) for a useful overview of research into multisensory processing.

content. Now let's say that I pick up the cup and hold it in my hands. It's natural for us to say that I both see and touch the cup. However, when we try to capture what conscious perceptual experience or experiences I have in this situation, there seem to be at least two possibilities. We might think that I have a visual experience that represents the cup and it's size and I have a distinct haptic experience that represents the cup and it's size. This is the Sense Modalist's characterization of perceptual experience in the situation in question. What will permit us to say that I have two co-conscious experiences in this case is if there are two representational contents in play: a visual representation of the cup and a haptic representation of the cup.

Alternatively, though, we might think that in cases where I receive information about the same property or the same object from more than one sensory receptor, then the result is—or at least may on occasion be—a single representation of that property/object. When my eyes are directed on the cup and I hold it, I have a conscious perceptual experience with visuo-haptic representational content (Bayne [2014], pp. 22–24). If at least some of the kinds of multisensory interactions that take place within the brain produce single multimodal representational contents of this kind, this would give us one reason to reject Sense Modalism. What reasons might there be for thinking that our conscious perceptual experiences can have multimodal content in this sense?

One basis for the claim has been the idea that the sensory systems integrate information. For example, de Vignemont suggests that, 'an experience is multimodal if it results from integrative binding' ([2014b], p. 133) , where:

> Integrative binding results from the fusion of sensory information that is redundant. For example, I hear Tim say "bonjour" and I see his lips moving, shaping the word "bonjour." Both modality-specific experiences carry information about the uttered word, which constitutes a common sensible. They can then be merged together into a unified content of the multimodal experience of the word. Because of the redundancy, the binding is so strong that the experiences melt into each other, so to speak. ([2014b], p. 130)

For de Vignemont, then, integrative binding produces multimodal experiences with unified contents.[4] We can, I think, understand this as the claim that integrative binding produces perceptual experiences with what I have called multimodal representational contents. What, though, is integrative binding? A little later de Vignemont explains that:

> [Integrative binding] is focused on a single property of the object. But we know that each sensory receptor sends noisy signals. Furthermore, the quality of the signals can be decreased by environmental conditions (poor light, for example, or noisy environment). It is thus important to have more than one source of information. Informational redundancy increases robustness and reliability. Thanks to integrative binding, the perceptual system can generate the best estimate of the property of the object by pondering and integrating the various sources of information. ([2014b], p. 132)

As we will see in more detail in the next section, integrative binding has all the hallmarks of models of multisensory integration: a particular kind of sensory interaction that has been proposed in the psychological literature. As with multisensory integration, integrative binding is a weighted multisensory interaction in which redundant information is merged together in order to

[4] For de Vignemont integrative binding does not exhaust the ways in which perceptual experience might count as multimodal. She proposes that a second interaction, which she calls additive binding, will also be sufficient for multimodality.

provide the most reliable estimate of a particular property of an object. What is more, in her explanation of integrative binding, de Vignemont refers her reader to an introduction to Bayesian accounts of perception that focuses on multisensory integration (Bennett *et al.* [2014]). Integrative binding can therefore, I think, be understood as a label for multisensory integration.

So, de Vignemont argues that experiences that are the product of multisensory integration will be multimodal. And, as we saw earlier, she takes multimodal experiences to be ones in which information has been merged together into unified contents. So, de Vignemont seems to be suggesting that experiences that are the product of multisensory integration will have what I have called multimodal perceptual contents.

What reasons do we have for thinking that integrative binding or multisensory integration takes place? De Vignemont does not give us any reasons for thinking that it does. But what makes her proposal particularly interesting is that in recent years there have been a number of empirical studies that have sought to determine whether the sensory systems integrate information in what is taken to be the optimal way according to the Bayesian approach to perception. And the results from these studies seem to suggest that our perceptual systems are optimal Bayesian decision makers.

In what follows, I aim to assess whether these empirical studies provide support for de Vignemont's claim that perceptual experiences can be multimodal. I'll focus on one study that examines the integration of visual and haptic information. My aim is to show that we should not assume that the sensory systems integrate information, and nor should we assume that any integration of sensory information that does take place sub-personally is in the service of multisensory perceptual experience.

In doing so, I do not set out to offer a defence of Sense Modalism. Indeed, doing so is beyond the scope of this paper, not least because there are many other ways of arguing for the multimodality of conscious perceptual experience. For example, it has been claimed that we perceive particular objects to bear a number of properties, each of which is perceptible through one sense modality only (Nudds [2014]; O'Callaghan [2014a]; Bayne [2014]; de Vignemont [2014b]). Casey O'Callaghan ([2017]) has argued that our perception of feature instances can be multimodal when aspects of the feature happen to be perceptible through different sense modalities on that particular occasion, say in cases of intermodal motion perception or intermodal rhythm perception. O'Callaghan ([2017]) has also argued that particular kinds of property that we perceive—for example, the flavour of mint—are only ever perceptible through the combined operations of more than one sense modality. These offer important lines of attack against Sense Modalism, and it's not clear how the Sense Modalist might successfully defend her position, other than by simply denying that we perceive the objects or properties in question.

But the plausibility of particular claims and the persuasiveness of certain lines of argument should not lead us to accept, uncritically, all proposals concerning the multimodality of perceptual experience. In what follows, therefore, I aim to assess whether some recent empirical work on multisensory integration provides any support for the philosophical claim that perceptual experience can have multimodal content, and hence provides us with a reason to reject Sense Modalism. To do this, I need first to say something more about multisensory integration.

3 Multisensory Processing and Maximum Likelihood Estimation

In recent years, a number of models have been proposed, each promising to provide a framework with which to understand many of the multisensory effects we have observed. For example, Sensory Combination describes interactions between sensory signals in different coordinate systems, or between sensory cues to different properties of the same object (Ernst and Bülthoff [2004], p. 163). Multisensory Integration, by contrast, in these cases picks out interactions between sensory cues to the same properties of the same object.[5]

While the different sensory systems process lots of information about different properties and objects, there is also some overlap. The sensory systems receive multiple cues to the same properties of the same objects. This redundancy of information is found within a sense, but also across the senses: for example, both the visual system and the auditory system process cues to spatial location. If the source of an audible sound is seen, then the sensory systems will process both visual and auditory information about the location of the sound source.

Sensory neurons have receptive fields. That is, each neuron will be activated by stimuli that fall within a particular region of space relative to the perceiver. A flash of light in front of you will evoke a response in the neurons whose receptive fields include the location of the light flash. If there was no neural noise, and the only visual stimulation corresponded to the flash of light, then only those neurons would be activated. But the presence of noise in the nervous system—spontaneous firing of neurons—means that some neurons whose receptive fields do not include the location of the light flash will also fire. The result is that sensory signals are unreliable and, at least somewhat, inaccurate. And so, the initial sensory estimates of a single property of a single object that are produced by different sensory systems can conflict. There might, for example, be a small discrepancy between the visual estimate of the location of the cup and the haptic estimate of the location of the cup.

According to proponents of multisensory integration, redundant sensory information is, at least much of the time, integrated to produce a single sensory estimate of a property. In this context, integration is a process of averaging information from two sources to produce the mean. Say, for example, that I measure the height of a cup twice, perhaps using two different methods. My measuring processes are imprecise, so I produce two different measurements of the height of the cup: one measurement of 4 cm and one measurement of 5 cm. If I integrate the two measurements, I average the two values to produce the mean measurement of the height. Assuming the two measurements are weighted equally, the result will be 4.5 cm.

The value of integration is that it can improve the reliability of the sensory estimates produced. According to a popular model for combining or integrating redundant information from more than one modality, Maximum Likelihood Estimation (MLE), our sensory systems integrate information optimally, as determined by the Bayesian approach to decision making. That is, they produce the most reliable—the most likely—estimate of the way things are in the world given the initial sensory information available. They do so by weighting each piece of information in the averaging process based on its relative reliability. The more reliable the information is, the more

[5] I am using the term 'multisensory integration' here to describe an interaction at the information-processing level. Note that the term is also used to describe the enhancement effects found in single cells in multisensory areas such as the superior colliculus of the brains of macaques and cats by Stein and Meredith ([1993]).

weight it is given proportionally, and so the greater contribution it makes to the average of the information from the two sources.

Take our toy example again and now let's say that I made more of an effort to measure the height of the cup accurately the first time round. I take that measurement to be more reliable than the second measurement. When I integrate my two measurements together I take the relative reliabilities of each measurement into account, assigning the first a weight of 0.75 and the second a weight of 0.25. In this case, I'll integrate the measurements to give a value of 4.25 cm.

MLE models the best or optimal way to combine redundant information because it produces a sensory estimate with the least variance and hence the greatest reliability. If the system is able to measure reliability well, improving reliability will improve accuracy. According to proponents of MLE, our sensory systems are optimal Bayesian decision-makers, integrating sensory information together in a way that heeds the relative reliability of each cue, and thereby producing sensory estimates that have less variance associated with them.[6]

It has been suggested that we can explain some of the multisensory effects we observe, in particular cross-modal illusions, in terms of multisensory integration (Briscoe [2016], p. 123; de Vignemont [2014b], p. 131; Bayne [2014], p. 21). The idea here is that multisensory integration as modelled by MLE will only be advantageous if the sensory cues that are integrated are in fact redundant. That is, it will only be advantageous if the sensory signals are derived from the same object or event. However, the suggestion goes, things can go awry—that is, cues that should *not* be integrated, are integrated—leading to cross-modal illusions. In the ventriloquism effect, for example, it's suggested that information about the location of what is heard is integrated with information about the location of what is seen, even though what it is that is seen and what it is that is heard are distinct entities.[7, 8]

While proponents of MLE are primarily engaged in giving a model of a sub-personal process, there are a number of reasons for thinking they might endorse the view that MLE models a process that produces multimodal perceptual contents.

First, MLE has been tested by asking participants to make reports about the stimuli with which they are presented. For example, participants are asked to judge which of two sequentially presented bi-modal stimuli is, for example, larger, or farther away. In giving a model of why it is that participants choose the stimulus they choose on each of the trials, the proponent of MLE seems to be giving an account of a personal-level phenomenon by offering a model of a sub-

[6] Generally speaking, Bayesian models are models of optimal decision-making in the face of uncertainty. Applied to perception, the Bayesian model treats the sensory systems as decision makers. The sensory systems produce estimates of properties in the environment on the basis of incoming cues. Within the Bayesian framework, the estimation of the property is the perceptual 'decision' that the system must make. According to the Bayesian account of perception, the sensory systems make use of prior information or assumptions about the way the world is likely to be with respect to a property to generate a more reliable specification of the property than would be provided by the incoming sensory information on it's own.

[7] See (Briscoe [2016]) for an explanation of a number of other cross-modal illusions in terms of MLE. See (de Vignemont [2014a], [2014b]) for the claim that the rubber hand illusion should be explained in terms of erroneous multisensory integration. Note that explanation of cross-modal illusions in terms of MLE doesn't require a commitment to the claim that the products of MLE are multimodal perceptual contents.

[8] It has been suggested that some cross-modal correspondences may be one of the factors that the sensory systems can use to determine when information from two senses concern the same object and this can be used to determine when information from those senses should be integrated (Parise and Spence [2009]; Chen and Spence [2017]), with some correspondences themselves arising from statistical regularities in the environment (Spence [2011]).

personal process (although of course, and as I shall argue, this need not be a perceptual phenomenon).

Moreover, in discussions of MLE we frequently find references to multimodal percepts. For example, Angelaki and colleagues introduce the MLE model by stating that 'a fundamental aspect of our sensory experience is that the information from different modalities is often seamlessly integrated into a unified percept' (Angelaki *et al.* [2009], p. 452). The sentiment is echoed by other leading figures in the MLE literature:

> To perceive the external environment our brain uses multiple sources of sensory information derived from several different modalities, including vision, touch and audition. All these different sources of information have to be efficiently merged to form a coherent and robust percept. (Ernst and Bülthoff [2004], p. 161)

And:

> We perceive our own body and the world surrounding us via multiple sources of sensory information derived from several modalities, including vision, touch and audition. To enable interactions with the environment this information has to converge into a coherent and unambiguous multimodal percept of the body and the world. (Ernst [2006], p. 105)

Of course, exactly what is meant by 'percept' in these quotes may be questioned. I won't pursue the question of whether we should take talk of multimodal percepts as indicative of commitment to the view that MLE produces multimodal perceptual contents here. Instead, I will take as my target the view put forward by de Vignemont ([2014b]), outlined in Section 2.

4 Experimental Support for MLE

The MLE model of integration can be used to make predictions about how perceivers will respond in situations in which two redundant but discrepant sensory signals are processed. This means the model can be, and has been, tested empirically. For example, an experiment by Ernst and Banks ([2002]) investigated how acquiring discrepant redundant visual and haptic information about the height of an object affected subjects' judgements about its height.

Subjects were asked to judge which of two stimuli, presented successively, was taller than the other. Both stimuli were bi-modal: that is, both stimuli had a haptic component and a visual component. On every trial participants both looked at and touched the stimulus at the same time.

However, the apparatus that Ernst and Banks used allowed them to manipulate the haptic and the visual components of the stimuli independently of one another. This meant that there could be differences between the visual and haptic height of the 'same' bi-modal stimulus. For example, on the same trial, the stimulus could have a visual height of 60 mm and a haptic height of 50 mm.

Of the two stimuli that participants had to compare on each trial, one had discrepant visual and haptic information about height, while the other had a uniform visual and haptic height. So, for example, the discrepant stimulus might have a visual height of 60 mm and a haptic height of 50 mm. By contrast, the uniform stimulus had the same visual height as haptic height, although that uniform visual-haptic height changed on a trial-by-trial basis.

What Ernst and Banks were interested in finding out was at what height the uniform stimulus had to be for a discrepant stimulus with a visual height of 60 mm and a haptic height of 50 mm to be judged to be the same size. This could be assessed by finding at what height the uniform stimulus had to be for subjects to judge it to be taller than the discrepant stimulus on 50% of trials. They called this the point of subjective equality (PSE).

Why were they interested in this? They asked subjects to perform the same task in four different conditions. In one condition the noise of the information was not experimentally manipulated. In the other three conditions they introduced visual noise at levels of 67%, 133% and 200% by increasing the blurriness of the visual stimulus.

This meant that they could look at the difference in the PSE in each condition, to see if differences in the reliabilities of the visual and haptic information relative to one another had any impact on how tall subjects judged the discrepant stimulus to be.

According to MLE, discrepant visual and haptic information about the height of a single object should be integrated to produce a perceptual decision about the height of the object. The combined estimate is the weighted average of the two unisensory estimates, with the weighting of information being proportional to their relative reliabilities (van Dam *et al*. [2014], pp. 210–212). If the reliability of the visual information is greater than the reliability of the haptic information, it will be given a greater weighting in the averaging process. As the reliability of the visual information drops relative to that of the haptic information, it will be given an increasingly smaller weighting in the averaging process.

Hence, the Bayesian model predicts that, as Ernst and Banks increased visual noise and made visual information less reliable, the PSE should change to reflect the change in relative reliabilities of visual and haptic information. Specifically, as the haptic information becomes more reliable relative to the visual information, the PSE will shift closer to the actual haptic height and further from the actual visual height of the discrepant stimulus.

This is precisely what Ernst and Banks found. When the discrepant stimulus had a visual height of 60 mm and a haptic height of 50 mm and there was no added visual noise, it was judged to be the same height as a uniform stimulus that was 58 mm tall (that is, the PSE occurred when the uniform stimulus was 58 mm tall). By comparison, when visual noise was at 200%, subjects judged the very same stimulus to be the same height as a uniform stimulus that was 52 mm tall.[9] So, subjects' judgements are as predicted by MLE.

Other experiments using different combinations of sense modalities, and cues to different kinds of properties have found that subjects perform as predicted by MLE (Ernst and Banks [2002]; Battaglia *et al.* [2003]; Alais and Burr [2004]). For example, similar two-interval forced choice (2I-FC) tasks were used by Battaglia *et al*. ([2003]) and by Alais and Burr ([2004]) to investigate subjects' judgements about spatial location when they are presented with discrepant visual and auditory information. They find similar results: subjects judge a uniform stimulus to be to the left of the discrepant stimulus on 50% of trials when the location of the uniform stimulus lies between

[9] Ernst and Banks were able to make more specific predictions about the PSE by establishing the reliability of the haptic modality on its own and the reliability of the visual modality on its own at the four different levels of visual noise used in their experiment. They were able to predict how the size of the comparison at the PSE would change as levels of visual noise were manipulated by using these reliabilities to work out the weighting of haptic and visual information at each level of visual noise and using those weights to work out an average of the actual visual and haptic heights of the discrepant stimulus. The results from the experiment were close to the predictions made by Ernst and Banks using MLE.

the visual and auditory locations of the discrepant stimulus, but closer to its visual location. The PSE differs, though, depending on the level of visual noise, in line with the predications of the Bayesian MLE model. Shams *et al*. ([2000]) tested numerosity judgements with sequences of auditory beeps and visual flashes. Bresciani *et al*. ([2006], [2008]) performed similar tests using sequences of tactile taps in combination with either visual flashes or auditory beeps.

5 Maximum Likelihood Estimation and Multimodal Perceptual Experiences

So, MLE seems to be backed up by results from psychophysics. What should we conclude from these predictive successes? MLE is a model of cue integration: it stipulates the way in which sensory information should be integrated, and so is a model of information processing. Ernst and Banks conclude, albeit somewhat cautiously, that 'the nervous system seems to combine visual and haptic information in a fashion similar to the MLE rule' (Ernst and Banks [2002], p. 431). And in an article about kinds of multisensory processing, Robert Briscoe avers that:

> A large and growing body of psychophysical evidence indicates that human perceivers do, in many cases, combine multisensory signals in the manner predicted by the MLE approach. (Briscoe [2016], p. 123)

But, de Vignemont's proposal about integrative binding and multimodality suggests we can conclude something further. As we have already seen, de Vignemont claims that:

> [. . .] an experience is multimodal if it results from integrative binding. What is bound is redundant information about the same property of the same object within the same spatial frame of reference. ([2014b], p. 133)

Redundancy of information is a feature of perception within and across the senses. Cues might be integrated within a sense when there is redundant sensory information available—say, when there are multiple cues to depth within vision. However, the experiments mentioned above indicate that participants behave as predicted when presented with redundant but discrepant sensory information in two different sensory systems. That is, we seem to have evidence of cue integration not only within a sense, but also across the senses.

On this basis it might be tempting to conclude that conscious perceptual experience can be multimodal, at least in so far as redundant information from multiple sources is integrated across sensory systems to produce single multimodal perceptual contents. We have evidence that supports MLE. If we combine this with de Vignemont's proposal that the processes that MLE models are processes that produce multisensory perceptual contents, then we might be tempted to conclude that, even if it is not always or even commonly the case, conscious perceptual experience can, on occasion, be multimodal. Hence Sense Modalism should be abandoned.

Should we accept that our conscious perceptual experiences can, at least some of the time, have multimodal perceptual contents, and so that Sense Modalism is mistaken?

The reductive assumptions underlying this reasoning might be of concern to us if we think there is good reason to retain a degree of autonomy between personal and sub-personal levels of description and explanation. However even if, for the sake of argument, we grant the reduction of personal level perceptual contents to sub-personal informational states, I don't think that the empirical evidence forces us to reject Sense Modalism.

In the remainder of this paper I will examine whether the psychophysical experiments designed to test MLE might also tell us about the multimodality or otherwise of the contents of perceptual experience. I will argue that, even when taken on it's own terms, the position adopted by de Vignemont is under-determined by the available evidence.

6 Integration or Modulation?

The first thing we should note is that results from MLE experiments are consistent with at least two different kinds of multisensory interaction. We can think of the sensory systems impacting each other in a number of ways. One possibility is that redundant information is integrated together to produce a single token multimodal representation of the property in question. Call this interaction Integration. Another possibility is that when redundant information is processed in two sensory systems, the 'integration' of information produces a combined estimate of the property, but one that is tokened multiple times in more than one sensory system. Call this interaction Modulation.[10]

O'Callaghan ([2017]) argues that cross-modal perceptual illusions such as the spatial ventriloquism effect and the rubber hand illusion can be explained in terms of mechanisms that recalibrate and coordinate responses across the senses ([2017], p. 158). Cross-modal perceptual illusions indicate that the sensory systems perform conflict resolution, but conflict resolution doesn't require that representations be shared between the senses. Resolution of conflict can be achieved using a simple set of 'if–then' rules, which don't produce multimodal representational contents.

For example, in the spatial ventriloquism effect, visual information about the location of a visible object conflicts with auditory information about the location of the source of a sound that is heard. When we are asked to report the location from which the sound came, we mislocate it in the direction of the visible object. But this need not require that the sensory systems have generated a single multimodal representation of an object, or of its location. We can equally explain the effect in the following way. The sensory systems encode a set of rules that specify pairs of outputs that should be produced in response to pairs of inputs. The visual system processes information about an object or event at location L_1. The auditory system processes information about an object or event at location L_5. According to the look-up table, the sensory systems should respond to this situation by producing a visual estimate of L_2 and an auditory estimate of L_2. As a result, the contents of the visual and of the auditory experiences will be different to the way they would have been had the subject been exposed to visual stimulation only, or auditory stimulation only. But, even though stimulation to one sensory system seems to shape our perceptual experiences associated with another sensory system (O'Callaghan [2017], p. 159), our conscious perceptual experiences can still each be associated with one and only one sense modality.

There might be other kinds of mechanism that could be classed as versions of Modulation. What matters for the Sense Modalist is that any Modulation mechanism will not produce a multimodal representational content.

[10] By focusing on these two potential ways the sensory systems might interact, I don't mean to suggest that these are the only two kinds of multisensory interaction that may occur.

The difference between Integration and Modulation is a difference in the products of the interaction. Integration functions to produce a single token perceptual representation of the property, while Modulation produces two or more token representations of the same type. Notice, though, that nothing else about the two mechanisms need be different. They can be subject to the same influences and sensitive to differences in the reliability of incoming information. In other words, they both can instantiate optimal Bayesian decision-making in the face of uncertainty, weighting the contributions of different sensory cues to the production of their estimate according to the relative reliability of each cue.

The problem for the opponent of Sense Modalism is that, as long as we are asking subjects to perform a 2I–FC task in which they must judge which of two bi-modal stimuli is the larger, Integration and Modulation will produce the same behavioural responses. For ease of exposition, I'll characterize the content of a perceptual experience of the height of the stimulus with a numerical value, '58 mm', say. According to Integration, when participants in the experiment have discrepant visual and haptic information about the size of the stimulus that is integrated to produce an estimate of it as being 58 mm tall, they have a visuo-haptic experience of the stimulus with a content '58 mm'. By contrast, according to Modulation, the result of the multisensory interaction will be a visual experience with a content '58 mm' and a haptic experience of the stimulus with a content '58 mm'. Whether the mechanism involved in the multisensory interaction is Modulation or Integration, subjects will still judge another stimulus, which they experience as being 64 mm tall, to be the larger.

We therefore cannot make substantial conclusions about the nature of the mechanisms involved in multisensory processing on the basis of the psychophysical evidence so far produced in support of MLE. The evidence supporting the Bayesian model is evidence of sensitivity to the noisiness of sensory signals, and the optimal weighting of that information in subsequent interactions. As long as a modulatory mechanism of the kind I've called Modulation can explain the behavioural results as well as Integration can, we don't have evidence that tells decisively in favour of the theory that sensory information is integrated to produce single token multimodal representational contents.

7 Perception or Judgement?

A further problem with the claim that evidence supporting the Bayesian model of MSI presents a challenge to Sense Modalism is that there are limits to what we can discover about perceptual experience using psychophysical experiments. Bayesian accounts of perception are concerned with perceptual 'decisions'—that is, with the production of sensory estimates on the basis of cues within perceptual processing. So, what the model predicts is what perceivers should have experienced when presented with the visuo-haptic stimuli. It assumes that changes in the reliability of the visual information should produce changes in the perceptual experiences of participants.

We know that as visual noise increases, subjects' judgements about the size of the visuo-haptic stimulus change. It appears that—as predicted by MLE—they are relying increasingly on the haptic information available, as its relative reliability improves. Judgements conform to the predictions made by the Bayesian model. But this means that what the experiments actually probe is how subjects' perceptual judgements change as the relative levels of noise in two sensory systems is manipulated. How, then, can we be sure that the integration occurs within perceptual processing, rather than post-perceptually, in the making of the judgement about which stimulus was larger?

Proponents of the Bayesian model have sought to address this question by looking for evidence of metameric behaviour. Perceptual colour metamers are colour stimuli that have different spectral compositions but are perceived by observers to be identical to one another in certain luminance conditions. What the exponent of the Bayesian model wants to find is evidence that subjects are similarly unable to discriminate between differently sized stimuli. The reasoning is as follows. If sensory information about a property is integrated to give a single multimodal representation of that property, then any stimulus with visual and haptic heights that average to 55 mm will be perceptually indistinguishable from any other stimuli whose visual and haptic heights also average to 55 mm, regardless of differences between the visual heights of the two stimuli and between the haptic heights of the two stimuli. These stimuli will be perceptual metamers of one another. So, Ernst says:

> [. . .] if two cues are totally fused at the perceptual level, the same percept will result whether they both indicate a medium value, or they differ radically from one another but average to a medium value. […] If cues are fused it should be obvious that also discrimination performance will be affected. Along the fused dimension [. . .] discrimination will remain possible. However, for stimuli not being along this dimension discrimination will drop and it will be impossible to discriminate stimuli that were averaged to exactly the same value – i.e., that are perceptual metamers. (Ernst [2006], pp. 114–5).

So, if sensory information is integrated to generate a single multimodal perceptual representation, then we might be able to find metameric behaviour. That is, we might find cases in which we fail to discriminate the perceptual metamers of a standard stimulus (stimuli that average to the same value as the standard) from the standard itself (Ernst [2006]).

Hillis and colleagues ([2002]) looked for metameric behaviour using a three-interval oddity task. Subjects were presented with three stimuli, two of which were identical, and were asked to pick the odd one out. They conducted two versions of the experiment. One was a within-modality experiment looking for evidence of the integration of binocular disparity and visual texture cues to slant. The other was a cross-modal experiment, looking for evidence of integration between visual and haptic information about size.

While there is (some) evidence of metameric behaviour for cues within a modality, they didn't find an effect in the cross-modal condition. What should we conclude from these results? The results do not look promising for the proponent of multimodality. Losing individual uni-modal sensory estimates seems to be a necessary consequence of integration within perceptual processing, and so the failure to find metameric behaviour might initially be thought to tell against the proponent of integration.

But Ernst ([2006]; see also van Dam *et al.* [2014]) suggests we can explain the differences in the within-vision and cross-modal cases here in terms of a distinction between 'mandatory or forced fusion'—in which information will always be integrated and access to the individual estimates is lost—and 'partial fusion'—in which information is integrated, but access to the individual uni-modal sensory estimates is retained. That is, Ernst argues that metameric behaviour is not even a necessary consequence of integration. Hence, we can explain the failure to find metameric behaviour in the cross-modal condition, without jettisoning the idea that sensory information is integrated.

What does it mean to say that information is integrated, but access to the individual uni-modal sensory estimate is retained?

If the sensory systems are to integrate information from more than one source, they must solve not one, but two 'multisensory' problems. They must establish *how* to integrate sensory information from more than one source—that is, they must determine the reliabilities of each sensory estimate if they are to integrate information in conformity with the principles of *MLE*. But first they must establish *when* sensory information should be integrated. That is, they must determine when sensory information is redundant: when information is about the same property of the same object. This is known in the literature under several different names, including the Causal Inference Problem (for example, Shams and Beierholm [2010]) or the Correspondence Problem (for example, van Dam *et al*. [2014]). The Bayesian model of cue integration solves the problem by postulating a coupling prior: a prior specifying the likelihood that information from two sources concerns the same property of the same object.[11]

Partial fusion has been postulated as the response made when the coupling prior that specifies that the likelihood that sensory information is redundant falls below a certain threshold. When there is such uncertainty about whether the two pieces of information have the same source, the system doesn't 'fully' integrate. That is, visual information influences auditory processing, and auditory information influences visual processing, but the outcome is *not* a single multimodal perceptual experience of the location of some entity. Rather, the multisensory interaction produces two distinct experiences with distinct representational contents.[12]

Partial fusion has thus commonly been invoked to explain results in which the presence of information in one sensory system influences the processing of sensory information in another sensory system (see, for example, Shams and Beierholm [2011]; Bayne [2014]). Take, for example, a typical study of spatial ventriloquism, in which subjects are asked to locate the source of the sound presented concurrently with a spatially discrepant visual stimulus. Subjects mislocate the auditory stimulus in the direction of the visual stimulus and the visual stimulus in the direction of the auditory stimulus, yet they do not typically (mis)locate the two stimuli in the same place. This can be explained in terms of partial fusion.

Notice that the notion of partial fusion seems to be very much akin to the mechanism Modulation, as outlined in the previous section. In both, what is perceived in one modality is changed as a result of what is perceived by means of the other modality, though both modalities are experienced. The difference between Modulation as introduced there and partial fusion as discussed here is that, in the latter the outcome of the interaction is not two token

[11] Other factors are also thought to influence how the perceptual systems solve the causal inference problem, including spatial and temporal coincidence, and semantic matching (Deroy *et al*. [2016], pp. 4–5).

[12] There is an alternative way in which we might understand the idea that information is integrated, but access to the individual uni-modal sensory estimates is retained. We might take the suggestion to be that in cases of partial fusion a process of integration produces a single sensory estimate of the property, and as a result the subject has an experience with multimodal representational content, but that, additionally, she has experiences with uni-modal representational contents that reflect the original sensory information available. When subjects are given a three-interval oddity task, their responses are based on the uni-modal experiences, and not the multimodal ones. Although I take this interpretation to fit with the description of partial fusion given by Ernst, I take it to be highly implausible that we might simultaneously have conflicting visual, haptic and visuo-haptic experiences.

representational contents that assign the *same* value to the same object, as it is in Modulation. Rather, partial fusion produces two token representational contents that assign different values to the objects represented in each modality.

Claiming that there is partial integration, understood in this way, in cases where there is uncertainty about whether information concerns the same property of the same object allows us to argue that subjects still have distinct (and discrepant) visual and haptic experiences of the bi-modal stimulus. This might be used to explain why Hillis and colleagues found no metameric behaviour in their cross-modal condition.

Yet, an account of the results in terms of partial integration does nothing to support the view that perceptual experiences have multimodal representational contents. Partial integration produces two perceptual representational contents, and not one multimodal content.

Moreover, the results found in the study do not give us reason to favour an explanation in terms of partial integration over the absence of perceptual integration. Subjects were able to discriminate bi-modal stimuli about as well as they were able to discriminate uni-modal stimuli presented on their own (in other words, with the same fineness of grain). As such, the claim that visual information influenced haptic processing at all, producing a haptic representation of the size of the (haptic) stimulus that was closer to the size of the visual stimulus, and vice versa seems to be unmotivated.

An alternative view, according to which integration is post-perceptual—that is, it is an effect we find when subjects make judgements about a property of an object in the world and their perceptual experiences have representational contents that are inconsistent with one another—seems equally well placed to explain both the results found in the 3-interval oddity task described in this section and the results from the two-interval forced-choice task discussed in Section 4.

8 Multisensory Integration for Action?

I've described two different multisensory mechanisms: Integration and Modulation. For the proponent of MLE to convince us to abandon Sense Modalism she needs to show that the kind of mechanism involved in the integration of sensory estimates is Integration. Yet the experimental results produced in support of MLE are consistent with both Integration and Modulation. Moreover, the evidence we have does not favour an interpretation in terms of the integration of information 'for perception' rather than 'for judgement'. Do we have any other reason for thinking that the interaction responsible for participants' responses is Integration rather than Modulation?

Ernst suggests that, where there is discrepant redundant sensory information available, that information must 'converge to a single multimodal percept' in order for us to be able to interact with our environment (Ernst [2006]). The thinking here seems to be as follows. The presence of noise in the nervous system means that information about the same property of the same object in different sensory systems will very likely differ, at least a little. There will be small discrepancies between the information that the visual system and the haptic system acquire about the location of a cup, say. But, what if I want to pick up the cup? I can't reach to two distinct places at the same time with the same hand, so I need to use one specification of location in order to act. How might I go about selecting sensory information in order to act? The optimal thing to do would be to integrate information from vision and touch about the location of the cup, so as to produce the most reliable, and hence the most accurate, estimate of its location.

There are a number of problems with this reasoning, principal amongst these being the assumption that any process of integration that could be required in order for us to act successfully must be in the service of perception. If we accept for the sake of argument that Integration is required for successful action, it's nevertheless possible that this integration process occurs independently from conscious perception.

Our ordinary, everyday assumption is that action is guided by conscious perceptual experience. On this assumption, the specification of a particular spatial property that guides action will be a component or part of the content of perceptual experience. However, the claim that action is guided by conscious perceptual experience has come under attack in recent years. Specifically, the assumption that the fine-grained details of motor actions requiring consistent representations of spatial properties are guided by conscious perceptual experience has been challenged. We have neurophysiological evidence from macaques with lesions in the dorsal or ventral streams (Ungerleider and Mishkin [1982]) and from neurological patients with visual agnosia and blindsight on the one hand or optic ataxia on the other hand (Goodale and Milner [1992]; Milner and Goodale [2006]), as well as behavioural evidence based on the responses of healthy human subjects to illusory stimulation which together provides support for a functional division between two streams of processing in the visual system. Based on this, it's been claimed that there is a functional division between processing in the dorsal stream 'for action' and processing in the ventral stream 'for perception'. Further evidence has been taken to support a division of labour between sensory processing for conscious perception and sensory processing for action in the auditory system (see, for example, Arnott *et al*. [2004]) and the somatosensory system (for example, Dijkerman and de Haan [2007]).

The distinction between sensory processing 'for perception' and 'for action' calls into question the idea that sensory information must be integrated so as to produce a conscious sensory experience with a single multimodal perceptual content in order for us to be able to perform world-directed bodily actions successfully. That is, even if our capacity to act depended on producing a weighted average of discrepant information from different sources, there may be room to argue that the process of producing the weighted average could be limited to the processing streams for action and therefore be unconscious. The idea that sensory information is integrated to guide motor action can be found already in (Stein [1992]), which suggests that the posterior parietal cortex is the locus of integration for motor activity.[13]

Related to this, Fisher and Pylyshyn contrasted pointing and verbal responses in a visuo-auditory localization task. That is, they contrasted mislocalization of spatially discrepant bi-modal stimuli in a task in which subjects merely gave a verbal report about where the stimuli were located with performance in a mislocalization task in which subjects were required to point to the location of the stimuli. They report that they, 'found a higher level of visual capture for motor performance than for apparent (cognitive) location of an auditory target in the presence of a visual distractor' (Fisher and Pylyshyn [1994], p. 95).

[13] See (Renier *et al*. [2009]) for empirical evidence that supports the claim that there is multisensory integration of sensory information processed within the action streams.

Of course, the possibility that some instances of integration occur within the stream 'for action' does not preclude the possibility that other instances of integration occur within the stream 'for perception'. For example, Fisher and Pylyshyn hypothesise that 'visual/auditory matching in motor performance behaves differently from matching for cognitive perception, implying that the "two visual systems" hypothesis may be better described as "two perceptual systems", and suggesting that at least two cross-modal localization modules may exist' (Fisher and Pylyshyn [1994], p. 95).

What is more, dual sensory processing system theories are controversial. But, we don't need to maintain that there is a functional division between different processing streams within each of the sensory systems in order to deny that the performance of successful motor actions shows that conscious perceptual experience is multimodal. This is because we need not accept that successful motor action requires Integration. We've already examined Modulation as an alternative way in which discrepancies in the information in different sensory systems might be resolved. There are also other ways in which a single specification of a spatial property could be reached or decided upon. For example, such a specification might be reached by deriving information about the relevant property from one sense only (Nudds [2014]). If I have discrepant visual and auditory information about the location of a bird and I want to point to it, I can base my pointing on the visual information only, or on the auditory information only. I don't have to integrate the information in order to perform the pointing action. It's simply not the case that discrepancies between sensory information must be resolved in order for us to act.

Contrary to the claims made by proponents of MLE, we shouldn't assume that our capacity to perform successful bodily actions in the face of discrepancies in sensory stimulation across the senses indicates that our conscious perceptual experiences have multimodal representational contents.

9 Conclusion

The empirical sciences continue to generate evidence about the operations of the sensory systems. Amongst other things, they show us that these systems interact with one another, and that these interactions result in behavioural effects.

One model of some of these multisensory interactions that has received a good deal of attention in recent years has been the Bayesian model of MLE, which models the integration of sensory information. This is a model at the sub-personal information processing level. De Vignemont ([2014b]) suggests, though, that MLE—or integrative binding—will produce unified multimodal contents at the personal level. The empirical evidence produced in support of MLE might, then, be taken as support for the claim that conscious perceptual experiences can be multimodal.

In this paper I have sought to argue that it would be a mistake to think that these empirical studies offer support for the claim that conscious perceptual experiences have, or can have, multimodal perceptual contents. Focussing on MLE across vision and touch, I have examined a particular study conducted by Ernst and Banks, designed to test their model of multisensory integration.

When it comes to multisensory integration across vision and touch, I have suggested that we have evidence that our judgements conform to the predictions made by the Bayesian MLE model, but this doesn't provide any conclusive proof that sensory information is integrated to produce a multimodal perceptual content.

This does not mean the empirical project of establishing that humans conform to the predictions made by the MLE model is not a valuable one. Neither does it mean that the perceptual systems do not weight sensory information in accordance with MLE. But it does mean that the empirical evidence produced in support of Ernst and Banks' model does not, by itself, occasion any kind of revision to a philosophical account of our ordinary view of conscious perceptual experience, namely Sense Modalism. To reject Sense Modalism, we must look to other proposals about how the sensory systems interact or other arguments for the multimodality of perceptual experience.

Acknowledgements

I would like to thank Matthew Nudds, audiences at the Stuttgart-Tübingen Research Forum and the Oxford Mind Work in Progress Seminar, and two anonymous reviewers for their insightful comments on earlier drafts of this paper. Funding for this research was provided by an Arts and Humanities Research Council grant for the 'Rethinking the Senses' project (AH/L007053/1).

Alisa Mandrigin
Division of Law and Philosophy and Division of Psychology
University of Stirling
Stirling, UK
alisa.mandrigin@stir.ac.uk

References

Alais, D. and Burr, D. [2004]: 'The Ventriloquist Effect Results From Near-Optimal Bimodal Integration', *Current Biology*, **14**, pp. 257–62.

Angelaki, D.E., Gu, Y. and DeAngelis, G.C. [2009]: 'Multisensory Integration: Psychophysics, Neurophysiology and Computation', *Current Opinion in Neurobiology*, **19**(4), pp. 452–8.

Arnott, S.R., Binns, M.A., Grady, C.L. and Alain, C. [2004]: 'Assessing the Auditory Dual-Pathway Model in Humans', *Neuroimage*, **22**(1), pp. 401–8.

Auvray, M. and Spence, C. [2008]: 'The Multisensory Perception of Flavour', *Consciousness and Cognition*, **17**, pp. 1016–31.

Battaglia, P.W., Jacobs, R.A. and Aslin, R.N. [2003]: 'Bayesian Integration of Visual and Auditory Signals for Spatial Localization', *Journal of the Optical Society of America A,* **20**(7), pp. 1391–7.

Bayne, T. [2014]: 'The Multisensory Nature of Perceptual Consciousness', in D. Bennett and C. Hill (*eds*) *Sensory Integration and the Unity of Consciousness,* Cambridge, MA.: Massachusetts Institute of Technology Press, pp. 15–36.

Bennett, D.J., Trommershäuser, J. and van Dam, L.C.J. [2014]: 'Bayesian Modeling of Perceiving: A Guide to Basic Principles', in D. Bennett & C. Hill (*eds*) *Sensory Integration and the Unity of*

Consciousness, Cambridge, MA.: Massachusetts Institute of Technology Press, pp. 3–13.
Bertelson, P. [1999]: 'Ventriloquism: A Case of Crossmodal Perceptual Grouping', in G. Aschersleben, T. Bachmann and J. Müsseler (*eds*) *Cognitive Contributions to the Perception of Spatial and Temporal Events*, Amsterdam: Elsevier Science, pp. 347–62.
Bresciani, J.-P., Dammeier, F. and Ernst, M.O. [2006]: 'Vision and Touch are Automatically Integrated for the Perception of Sequences of Events', *Journal of Vision,* **6**(5), pp. 554–64.
Bresciani, J.-P., Dammeier, F. and Ernst, M.O. [2008]: 'Tri-Modal Integration of Visual, Tactile and Auditory Signals for the Perception of Sequences of Events', *Brain Research Bulletin,* **75**, pp. 753–60.
Briscoe, R.E. [2016]: 'Multisensory Processing and Perceptual Consciousness: Part I', *Philosophy Compass,* **11**(2), pp. 121–33.
Calvert, G.A., Spence, C. and Stein, B.E. [2004]: *The Handbook of Multisensory Processes*, Cambridge, MA.: Massachusetts Institute of Technology Press.
Chen, Y.-C. and Spence, C. [2017]: 'Assessing the Role of the 'Unity Assumption' on Multisensory Integration: A Review', *Frontiers in Psychology*, **8**, p. 445. doi: 10.3389/fpsyg.2017.00445
van Dam, L.C.J., Parise, C.V. and Ernst, M.O. [2014]: 'Modelling Multisensory Integration'. In D. Bennett and C. Hill (*eds*) *Sensory Integration and the Unity of Consciousness*, Cambridge, MA.: Massachusetts Institute of Technology Press, pp. 209–29.
Deroy, O., Spence, C. and Noppeney, U. [2016]: 'Metacognition in Multisensory Perception', *Trends in Cognitive Sciences*, **1601**, pp. 1–12. DOI: 10.1016/j.tics.2016.08.006
Dijkerman, H.C. and de Haan, E.H.F. [2007]: 'Somatosensory Processes Subserving Perception and Action', *Behavioral and Brain Sciences*, **30**(2), pp. 189–201.
Ernst, M.O. [2006]: 'A Bayesian View on Multimodal Cue Integration', in G. Knoblich, I. Thornton, M. Grosjean, and M. Shiffrar (*eds*) *Human Body Perception from the Inside Out,* Oxford: Oxford University Press, pp. 105–31.
Ernst, M.O. and Banks, M.S. [2002]: 'Humans Integrate Visual and Haptic Information in a Statistically Optimal Fashion', *Nature,* **415**, pp. 429–33.
Ernst, M.O. and Bülthoff, H.H. [2004]: 'Merging the Senses into a Robust Percept', *Trends in Cognitive Sciences*, **8**(4), pp. 162–9.
Fisher, B.D. and Pylyshyn, Z. [1994]: 'The Cognitive Architecture of Bimodal Event Perception: A Commentary and Addendum to Radeau', *Current Psychology of Cognition*, **13**(1), pp. 92–6.
Gallace, A. and Spence, C. [2006]: 'Multisensory Synesthetic Interactions in the Speeded Classification of Visual Size', *Perception and Psychophysics*, **68**, pp. 1191–203.
Goodale, M.A. and Milner, A.D. [1992]: 'Separate Visual Pathways for Perception and Action', *Trends in Neurosciences*, **15**(1), pp. 20–5.
Hillis, J.M., Ernst, M.O, Banks, M.S. and Landy, M.S. [2002]: 'Combining Sensory Information: Mandatory Fusion Within, But Not Between, Senses', *Science*, **298**, pp. 1627–30.
Jousmaki, V. and Hari, R. [1998]: 'Parchment-Skin Illusion: Sound-Biased Touch', *Current Biology,* **8**, p. R190.
McGurk, H. and MacDonald, J. [1976]: 'Hearing Lips and Seeing Voices', *Nature*, **264**, pp. 746–8.
Milner, A. D. and Goodale, M. A. [2006]: *The Visual Brain in Action*. Oxford: Oxford University Press.

Nudds, M. [2014]: 'Is Audio-Visual Perception Cross-Modal or Amodal?', in D. Stokes, M. Matthen and S. Biggs (*eds*) *Perception and Its Modalities*. Oxford: Oxford University Press, pp. 166–88.

O'Callaghan, C. [2014a]: 'Intermodal Binding Awareness', in D.J. Bennett and C.S. Hill (*eds*) *Sensory Integration and the Unity of Consciousness*. Cambridge, MA.: Massachusetts Institute of Technology Press, pp. 73–103.

O'Callaghan, C. [2014b]: 'Not All Perceptual Experience is Modality Specific', in D. Stokes, M. Matthen, and S. Biggs (*eds*) *Perception and Its Modalities*. Oxford: Oxford University Press, pp. 133–65.

O'Callaghan, C. [2017]: 'Grades of Multisensory Awareness', *Mind and Language,* **32**(2), pp. 155–181.

Parise, C. and Spence, C. [2009]: 'When Birds of a Feather Flock Together: Synesthetic Correspondences Modulate Audiovisual Integration in Nonsynesthetes', *PLoS ONE*, **4**, pp. e5664. doi: 10.1371/journal.pone.0005664

Renier, L. A., Anurova, I., De Volder, A. G., Carlson, S., VanMeter, J. and Rauschecker, J. P. [2009]: 'Multisensory Integration of Sounds and Vibrotactile Stimuli in Processing Streams for "What" and "Where"', *The Journal of Neuroscience*, **29**(35), pp. 10950–60.

Shams, L. and Beierholm, U.R. [2010]: 'Causal Inference in Perception', *Trends in Cognitive Sciences*, **14**, pp. 425–32.

Shams, L. and Beierholm, U. [2011]: 'Humans' Multisensory Perception, from Integration to Segregation, Follows Bayesian Inference', in J. Trommershäuser, K. Kording and M.S. Landy (*eds*) *Sensory Cue Integration*. Oxford: Oxford University Press, pp. 251–62.

Shams, L., Kamitani, Y. and Shimojo, S. [2000]: 'What You See is What You Hear', *Nature*, **408**(6814), p. 788.

Smith, B.C. [2015]: 'The Chemical Senses', in M. Matthen (*ed*) *The Oxford Handbook of Philosophy of Perception*. Oxford: Oxford University Press, pp. 314-52.

Spence, C. and Bayne, T. [2013]: 'Is Consciousness Multisensory?', in D. Stokes, M. Matthen, and S. Biggs (*eds*) *Perception and its Modalities*. Oxford: Oxford University Press, pp. 95–132.

Spence, C. [2011]: 'Crossmodal Correspondences: A Tutorial Review', *Attention, Perception, & Psychophysics*, **7**(4), pp. 971–95.

Stein, B. [1992]: 'The Representation of Egocentric Space in the Posterior Parietal Cortex', *Behavioral and Brain Sciences*, **15**(4), pp. 691–700.

Stein, B.E. and Meredith, M.A. [1993]: *The Merging of the Senses*. Cambridge, MA.: Massachusetts Institute of Technology Press.

Ungerleider, L.G. and Mishkin, M. [1982]: 'Two Cortical Visual Systems', in D.J. Ingle, M.A. Goodale, and R.J.W. Mansfield (*eds*) *Analysis of Visual Behavior,* Cambridge, MA: Massachusetts Institute of Technology Press.

de Vignemont, F. [2014a]: 'A Multimodal Conception of Bodily Awareness', *Mind*, **123**(492), pp. 989–1020.

de Vignemont, F. [2014b]: 'Multimodal Unity and Multimodal Binding', in D.J. Bennett and C.S. Hill (*eds*) *Sensory Integration and the Unity of Consciousness*, Cambridge, MA.: Massachusetts Institute of Technology Press, pp. 125–50.

Welch, R.B. and Warren, D.H. [1980]: 'Immediate Perceptual Response to Intersensory Discrepancy', *Psychological Bulletin*, **88**, pp. 638–67.