Is audio-visual perception 'amodal' or 'crossmodal'?

Matthew Nudds

The senses are modal in the following ways. The different senses – or, at least, the senses of vision, touch, and hearing[1] – each function to enable us to perceive objects and their features, and each can operate independently of the others. We can see something without hearing or touching it, hear something without seeing or touching it, and so on. Each sense modality is, therefore, (relatively) functionally independent of the others. In addition, each sense modality enables the perception of a range of modality specific objects or features – objects or features that can only be perceived with that particular sense. We can only see colours, only hear sounds and their features, only feel heat, and so on.

Although the senses are (relatively) functionally independent of each other, they do not, for the most part, operate independently of each other. Our perceptual experience at any time is the result of the simultaneous operation of all of our senses, and many of the things we perceive we perceive with more than one sense simultaneously. We often perceive the same particular thing, and the same features of that particular thing, with more than one sense. If you look at a coin you hold in your hand, you both see and feel the coin, and you can both see and feel its shape. When you drop the coin to the floor, you can both see and hear it strike the floor, and both see and hear when and where it strikes the floor. So although the senses are modal, perception is often multi-sensory.

Multi-sensory perception[2] might be supposed to be simply the combined operation of each of the individual senses. That is, the multi-sensory perception of something might be supposed to be the combination of what would be perceived by each sense operating independently of the others; and the awareness we have of something with features perceived with different senses to consist in the post-perceptual combination of what is perceived with each sense individually.

This picture of multi-sensory perception cannot be right. We can perceive the same properties of a particular object using more than one sense modality. So the different sensory processing streams that constitute these different sense modalities process information about the same features of the same things. For example, both hearing and vision process spatial information about the same events and objects. The fact that this information comes from the same source object does not guarantee that it will match across the different processing streams. A particular stream may be affected by noise, or by conditions that prevent its

[1] The following is not obviously true of the senses of smell and taste. It is arguable that smell and taste enable us to perceive only smells and tastes rather than objects that have smells and tastes, and they may not be functionally independent of each other.

[2] I use the term 'multi-sensory perception' for any perception by an individual involving more than one sense modality, irrespective of whether there is any integration of information across senses; 'multi-modal perception' is perception of something involving a multi-modal sensory process that integrates or combines information across different sensory systems. We should allow for the possibility that not all multi-sensory perception is multi-modal perception in this sense.

optimal operation, so that although the spatial information in the two streams derives from the same distall object it may be inconsistent across streams, or be much less accurate in one stream than in the other.

The function of perception is, at least in part, to guide action. For example, the spatial content of the perceptual experiences of an object determines the spatial properties of object directed actions. In order for perception to perform its action-guiding function, whenever the same thing is perceived with more than one sense, some process is required to eliminate inconsistencies. Suppose that you can both hear and see some object that you want to reach. If the spatial information concerning the location of the object is different in vision and audition, then in order to determine the trajectory and endpoint of your reaching the inconsistency must be eliminated. Either spatial information from one sensory modality must be selected over the other, or information from both senses must be integrated. So, the fact that we act on particular things that we perceive with more than one sense requires that there be processes to select or integrate information across senses. If the spatial information concerning the location of the object is more accurate in one stream than in the other, then these processes must either select the most accurate information, or combine or integrate information in some way to enhance its accuracy. It is only by doing this that reaching is likely to be successful.

Multi-sensory perception cannot, therefore, simply consist in the combined operation of each of the individual senses: some multi-sensory perception involves multi-modal perceptual processes. The fact that the same objects, and the same features of those objects, can be simultaneously perceived with more than one sense means that there must be inter-sensory connections between different sensory systems, and inter-sensory integration of information across senses. This argument appeals to the action-guiding function of perception, but a similar argument could be made by appeal to the fact that function of perception is to produce accurate or veridical perceptual states. What is the nature of these inter-sensory connections and integration of information and what do they tell us about perception and the nature of the senses? In this chapter I try to shed light on these questions by focusing on audio-visual interactions.[3]

In auditory perception we hear things in virtue of hearing the sounds they produce. Sounds are individual things that can instantiate a range of different acoustic properties, such as loudness and pitch. These properties are modality specific. It is possible for us to hear a number of distinct sounds at the same time – the sound of a bird outside, the buzzing sound made by the computer, and so on – each of which instantiate a range of acoustic properties.

Sounds are distinct from their sources – the things that produce them.[4] The source of a sound is often a material object, something that instantiates a range of non-acoustic properties

[3] There are similar kinds of interactions to those I discuss involving vision and touch, and much of what I say generalises to them (interactions involving flavour, taste, and smell perception appear to be different).

[4] There are a number of different accounts of the nature of sounds. See, for example, Pasnau (1999),

such as shape, size, and colour. Since sounds and material objects are not indiscernible, sounds are not identical with material objects. A sound is produced only when something happens to a material object – when an event of some kind occurs – so sounds are produced by events. In most cases, our ordinary ways of individuating the events that produce sounds distinguishes them from the sounds they produce: a particular sound may have been produced by the breaking of the glass,[5] but the breaking of the glass produces a number of distinct sounds, so the particular sound is not identical to the breaking of the glass. So sounds are not identical to the events that produce them, at least not as those events are ordinarily individuated.[6]

It would be a mistake to think that because we hear the sources of sounds in virtue of hearing the sounds that they produce, the content of auditory perception is restricted to sounds and their features. Berkeley held this kind of view: "when I hear a coach driving along the streets, all I immediately perceive is the sound… I am said to 'hear the coach'… [but] *in truth and strictness nothing can be heard but the sound*." His reasons for thinking this derive from his empiricism: it is not possible to explain how auditory perception could be the perception of anything other than sounds within that empiricist framework. If we accept Berkeley's restriction, crossmodal interactions involving auditory perception are inexplicable: those interactions involve the integration of information about the same objects and features perceived with more than one sense. On a Berkeleian view of sounds, sounds are distinct from the material objects we see, so we never see and hear the same objects and features. Furthermore, the integration of information from vision and hearing would have to involve the integration of information from different objects and different features: information about the sounds we hear – sounds which are distinct from material objects – somehow being integrated with information about the features of material objects perceived with other senses.

We should reject Berkeley's restriction. Once we do so, we can allow that the purpose or function of auditory perception is the perception of the sources of sounds, and not simply the sounds they produce. If that's right, then auditory perception has the function of representing the sources of sounds. We perceive sound sources by perceiving the sounds they make, so we perceive sounds and their sources,[7] but we shouldn't think that sounds somehow get in the way of our perceiving their sources, or cut us off from them: quite the opposite – they put us in touch with their sources. A proper defence of these claims requires an account of auditory perception – of how we perceive sounds and their sources. I don't have space for such an

---

Casati and Dokic (2005), O'Callaghan (2007), Matthen (2010), and the papers in O'Callaghan and Nudds (2010). My discussion in this chapter is, I think, neutral between these different accounts.

[5] An event that begins with my hitting the glass and ends with pieces of glass at rest on the floor.

[6] It might be suggested that we can think of the sound of the breaking of the glass as a single sound, but our doing so is a result of hearing a sequence of individual sounds as grouped or connected. It's always possible to specify a source event in terms of the sound it produces (a ringing in terms of a ringing sound, a scratching in terms of a scratching sound). The claim that we hear the sources of sounds is not simply the claim that we hear such 'sounding' events. An argument along these lines is defended in more detail by O'Callaghan (2007, 21ff).

[7] Brian Loar (1996, 144ff.) similarly argues that olfaction represents smells and their sources.

account here.[8] In what follows I will assume that auditory perception represents both sounds and their sources, and hence that multi-modal perception involving auditory perception involves the multi-modal perception of the sources of sounds, not of sounds; and that inter-sensory integration involves the integration of information about the features of sound sources.

*

We can perceive the same particular thing with more than one sense. What is involved in perceiving particular things in vision and in audition?

In the case of vision, we see particular objects. We do so in virtue of our visual experience representing them as such. The visual process in virtue of which we see objects is relatively well understood. It begins with information extracted from the patterns of illumination detected by the retina. Early visual processing involves a number of distinct retinotopic feature maps that operate in parallel to analyse and extract information about different features of objects. Features from different maps that correspond to the same distal object must be grouped or 'bound' together into states that correspond to that object. The result of binding is that a conjunction of features that are likely to be features of the same distal object are grouped together. One influential account of binding hypothesises that features are bound together on the basis of their spatial location.[9]

This kind of feature binding is necessary for object representation, but to represent something as an object requires more than just representing a conjunction of features at a particular place. It requires representing those features as features of something that is cohesive, bounded, and spatio-temporally continuous.[10] To say that something is represented as cohesive, bounded, and spatio-temporally continuous is to say that it is represented as having parts that belong together as parts of the same object (the parts are all connected, and move together) and as distinct from other objects (so they won't merge with other objects they come into contact with), and as maintaining its identity over time and through changes in location. To be a representation as of an object a group of features must have an identity that is not just that of the conjunction of features.

One way to think of representations of objects is in terms what Kahneman and Triesman call 'object files'. Object files are representations that maintain the identity of an individual object through changes in its features. When an object is perceived, information about it is placed into a file. This information might include information about its features – its location, shape, colour, and so on. As more information about the object accumulates over time – information about, for example, the kind of object it is – it is added to the file. If the object moves or changes, then the object file is updated to reflect this.

---

[8] See Nudds (forthcoming).
[9] Treisman 1998.
[10] See Matthen (2005, ch.12) for a discussion of the idea that features are bound to particulars; see Burge (2010, ch.10) for a recent discussion of the necessary conditions for the visual representation of objects. Both contain further references to relevant empirical literature.

Evidence that visual information is grouped into object files comes from what is known as an *object specific preview benefit*.[11] This is demonstrated with an experiment in which two objects are presented on a screen. Letters are briefly displayed on each object, and the objects then move to a new position. When the objects stop, a letter is displayed on one of them. The subject's task is to name the letter. In 'same object' trials, the same letter is displayed on the object before and after it moves; in 'different object' trials, the letter that was displayed on the other object is displayed; and in a third set of trials a novel letter is displayed. Subjects are quicker at naming the letter in the same object trials than in the other trials. This can be explained on the assumption that information about objects is stored in a file.

In the case of audition, we hear particular sound sources, and we can hear a number of distinct sound sources simultaneously. The auditory process in virtue of which we hear sound sources is less well understood than the visual process in virtue of which we see objects. At any time the sound waves that reach the ear will have been produced by any number of distinct sound sources. In order to perceive individual sources, the auditory system must organise the different frequency components that make up the sound wave into groups (or streams) that normally correspond to distinct environmental sources,[12] with the result that grouped frequency components are experienced as a single sound. These frequency component groups or streams correspond to, and carry information about, the particular things in the environment that produced them. As well as organising frequency components into groups at a time, the auditory system groups sequences of frequency components over time in ways that correspond to their sources, with the result that sequences of sounds are experienced as belonging together. A sequence of sounds produced by a single source is normally experienced as having been produced by a single source; that is, a sound at one time is experienced as having been produced by the same source as sounds experienced earlier.[13] So in auditory processing there are states that correspond to and carry information about particular sound sources.

It is not clear what auditory experience represents the sources of sounds as: in particular, it is not clear whether it represents the sources of sounds as particular objects, or as events of certain kinds. In many cases the sources of sounds are particular objects, but sounds are normally only produced by something happening to a particular object. For example, sounds are only produced by a metal bar when it is struck or otherwise caused to vibrate. So does auditory perception represent the sources of sounds as the particular objects which produce sounds when something happens to them, or does it represent them as the sound producing events that happen to particular objects?

I suggested that to represent something as an object requires that it be represented as having properties constitutive of being an object: as cohesive, bounded, and spatio-temporally continuous. If we think that auditory perception cannot represent those kinds of properties

[11] See Palmer (1999, sec. 11.2.6) for a summary of the relevant experiments and of what they show, and further references.
[12] See Bregman 1994, ch.3, and Nudds (2010) for further discussion.
[13] See Bregman 1994, chs. 2 and 4.

then we might conclude that it does not represent the sources of sounds as objects, but merely as events happening to objects. For example, auditory perception doesn't represent the volumetric shape of sound sources, so if representing something as bounded and cohesive requires representing its volumetric shape, then auditory experience doesn't represent anything as bounded and cohesive. If that's right then we might doubt that auditory perception represents sound sources as objects.

We should be careful, however, not to rule out the possibility that auditory perception represents objects simply on the grounds that it lacks something that is distinctively visual. (Representing shape may only be required for representing something as an object when the input is a two-dimensional retinal array.) We can find auditory analogues of the properties that are constitutive of visual object representation. Particular sound sources are perceived as a consequence of the – in many cases, non-spatial[14] – way that auditory perception 'segments' the auditory scene; distinct sounds are grouped together in virtue of having the same source, so auditory perception is able to track sound sources over time and through changes; auditory perception is sensitive to whether a sound source maintains its cohesiveness over time – think of the difference between hearing a bottle drop to the floor and bounce, and hearing it break – and, since many sounds are such that their nature is partly determined by the structure – volume, shape, and material construction – of the object that produced them, many sounds are such that they could normally only have been produced by a single, cohesive, object. It would seem, then, that auditory perception can track the kinds of properties that are constitutive of being an object. So it's not implausible to suggest that auditory perception represents the sources of sounds as having properties that are constitutive of being an object. Furthermore, there is some evidence that information about sound sources is bound together into a representation – an 'auditory object' file – that functions in auditory perception in a way that is similar to the way object files function in visual perception.[15]

Both visual and auditory perception represent particular things and their features. Visual perception represents objects as such, auditory perception represents the sources of sounds, if not as objects, then as events.[16] In both cases, we can think of information about particular things as stored in 'object' files, allowing that auditory object files may have identity

---

[14] The role of spatial properties in auditory grouping is not straightforward. In vision, features are spatially 'indexed' and two features with the same spatial properties may be bound together. In audition features are not spatially indexed, but acoustic features may be bound together because they share non-spatial cues that indicate that they were produced at the same location. Sharing such spatial cues is, however, not necessary for acoustic features to be grouped. See Nudds 2009, pp. 78-83.

[15] Zmigrod and B. Hommel, 2009. This is not an area that has been much investigated. One reason for this is that in some discussions auditory objects are supposed to be sounds, rather than the sources of sounds, and so evidence is sought that there are object files that represent sounds and their features. Given what I have been arguing about sounds in relation to their sources what we want is evidence that there are object files that represent the *sources* of sounds and their features rather than sounds and their features.

[16] Although I think the case for saying that auditory perception represents sound sources as objects, the substance of the following discussion is not much affected if it turns out that auditory perception represents the sources of sounds as events.

conditions similar to those of events rather than of objects (in what follows, I will call these 'visual-object' representations and 'auditory-object' representations). Given this, how should we understand multi-sensory perception and the inter-sensory integration of information?

*

I argued that integration of information across different senses is necessary given that the same particular things, and the same features of those things, can be perceived with more than one sense modality. What does it mean to say that information is integrated? On the face of it, the integration of information required by multi-sensory perception is consistent with two different models: the 'crossmodal' model, and the 'amodal' model.

According to the amodal model, there are a number of distinct low-level processing streams corresponding to each of the different sensory modalities, but these processing streams lose their distinctness at higher levels. At lower levels, information in each stream may be grouped together in ways that correspond to the distal object from which it comes – that is, into visual-object and auditory-object representations – but at higher levels information from the distinct processing streams that corresponds to the same distal object is combined into a single 'amodal' representation of that object. Since the distinct low-level visual- and auditory-object representations will contain information concerning some of the same properties of an object, there are mechanisms which combine or integrate this information in an optimal way that maximises accuracy and resolves any inconsistencies. According to this model, the perceptual system represents objects amodally: a number of initially distinct processing streams combine to produce a single amodal representation of an object, that represents it as having features – such as spatial and temporal features – that may have been perceived with more than one sense modality, as well as features – such as colour – that are modality specific. A single amodal object representation may represent an object as shaped, coloured, and as the source of a sound. It follows that the same kind of amodal-object representation plays a role in explaining our perceptual awareness of particular things perceived with any of the sense modalities. Both our visual perception of an object and our auditory perception of a sound source is explained by appeal to the same kind of amodal object representation.

According to the crossmodal model, each sense modality that represents particular objects does so by means of modality specific object representations. There are a number of distinct processing streams corresponding to the different sensory modalities. In each stream there are representations of particular objects and their features – visual-object representations, auditory-object representations, and so on. In addition there are crossmodal connections that function to modulate the information within each of the processing streams in the light of information in the other streams in such a way as to maximise accuracy and consistency of these distinct object representations. The same distal object may be represented simultaneously by means of distinct object representations in two or more sensory modalities; when that happens the crossmodal mechanisms operate to ensure that the features an object is represented as having – in particular the spatial and temporal features – are consistent

across representations in the different sensory modalities. The modality-specific object representations are modulated, but remain distinct. Distinct kinds of modality-specific object representation play a role in explaining our perceptual awareness of things perceived with each of the sensory modalities. Our visual perception of an object is explained in terms of a modality specific visual-object representation of an object, our auditory perception of a sound source is explained in terms of a modality specific auditory-object representation of an object, and so on.

The difference between these two models is not in the existence of inter-sensory connections or the inter-sensory integration of information. Both models involve inter-sensory integration. The difference is in the effects of this integration: whether it results in a single amodal object representation or instead modulates several distinct modality specific object representations. In both cases a single distal or environmental object will be represented, but in one case it will be represented by means of an amodal object representation and in the other by two or more modality specific object representations.

According to the crossmodal model, each sense modality represents particular objects. The visual system represents particular objects, and the auditory system represents particular objects, and there are principles that ensure that when the same distal object is represented both visually and auditorily, information about the non-modality specific properties of these objects is integrated. The result of the integration is that, with respect to features that can be perceived with more than one sense, the visual-object representation and the auditory-object representation represent the object as having the same features, e.g., as being at the same location, occurring at the same time, and so on, but the visual-object representation of the object will also represent it as having features that are specific to vision, and the auditory-object representation of the object will also represent it as having features specific to audition. Perception is *crossmodal* in the sense that information from other senses contributes to and helps to determine what is represented in any particular sense modality. But it is not *amodal* because there are distinct object-representations of the same distal or environmental object in each of the sense modalities, and these distinct sense-specific object-representations explain our perceptual awareness of objects perceived with each of the senses.

According to the amodal model, rather than distinct sensory modalities each representing the same particular object, information from distinct sensory modalities is integrated into a single amodal representation of a particular object. The result is that there is a single amodal representation of an object that represents the object as having both features that can be perceived with more than one sense, and features that can only be perceived with one sense.

Of course, the existence of amodal object representations doesn't rule out the existence of modality specific object-representations, and a system that produces amodal object representations could do so by combining modally specific object representations.[17] So the

[17] It wouldn't necessarily operate in this way. If there are amodal representations then it's a further question what the mechanism are that produce them – are there modal representations that are combined, or simply mechanisms that produce amodal representations?

difference between the two models is not in whether there are modality specific object-representations. The amodal model could accept that there are. Rather, the difference concerns the existence of amodal object representations and the explanation of our perceptual awareness of distal or environmental objects: according to the amodal model, both our visual and our auditory perception of particular objects is explained by appeal to the same kind of amodal object representation, and according to the crossmodal model our visual and our auditory perception of distal or environmental objects – even when it is the same object perceived simultaneously with two senses – is explained by appeal to distinct – modality specific – object representations of the object.

Another way to think of the difference between the two models is in terms of the consequences inter-sensory interactions have for the way distal or environmental objects are perceptually represented. Do inter-sensory interactions result in objects perceptually represented by means of different kinds of (coordinated) modality specific object representations, or by means of a single type of amodal object representation? If the former, then our perceptual awareness of objects is explained in terms of modality specific object representations, if the latter then our perceptual awareness of objects is explained in terms of amodal object representations. These two different explanations have consequences for both the account we give of the veridicality conditions of experiences of objects, and for the account we give of object representation more generally.

If the crossmodal model is correct, then senses can be said to be fundamentally functionally independent of each other.[18] According to this model, when an object is perceived with a single sense modality our perception of it is explained in terms of a modality specific object representation produced by a process whose operation is fundamentally independent of the operation of the other senses. We only need to consider other senses when an object is perceived with more than one sense, and then only in order to understand how what is perceived in one sense is modulated by what is perceived in the other. If the amodal model is correct, the senses are not fundamentally functionally independent of one another. The operation of one sense modality cannot be understood in isolation from the other sense modalities, and we need to consider the other senses even when an object is perceived with only one sense. Even when an object is perceived with a single sense modality, our perception of it is explained in terms of an amodal representation produced by a process which is in part common across sense modalities, and whose operation is not independent of

[18] The very existence of inter-sensory interactions might be thought to undermine this claim. Fodor's original characterization of modularity (1983) views modules as encapsulated, with no exchange of information between them. Inter-sensory interactions might therefore be thought to show that the senses are not encapsulated, hence not modular. That conclusion, however, is too quick. The organization of the brain is such that there are significant connections between functionally specialized systems that we have good reason think are functionally independent (see Shallice (1988 ch.11) for a discussion of this point). Two things follow. First, Fodor's characterization of a module needs to be amended – in particular to allow *some* exchange of information – if it is to be of use in neuropsychological explanation; second, the functional organization that crossmodal model describes can be viewed as describing functionally specialized, independent, sense modalities. For more detailed discussion of these issues, see Nudds (2011).

them. So, for example, the explanation of our auditory awareness of the sources of sounds will appeal to the same representational capacity that we appeal to in order to explain our visual awareness of objects.

*

I have sketched these two models in a very general way; there are other, more complicated, possibilities that I am ignoring.[19] I am ignoring them because I am interested in the explanation of our perceptual awareness of particular objects: is our perceptual awareness of objects explained in terms of amodal representations of objects, or only ever in terms of modality specific representations of objects? The answer to this question has consequences for the veridicality conditions of perceptual experiences. If perceptual awareness of objects is explained in terms of amodal representations of objects, then when the same thing is both seen and heard it will be represented by means of a representation of it as having both visual and auditory features. For example, suppose you hear someone you can see speaking: you hear speaking and you see lip movements. Your perceptual experience will represent the person you see as the source of the sounds you hear. If the person you see is not the source of the sounds, then the representation, and hence your perceptual experience, is non-veridical. If, on the other hand, perceptual awareness of objects is explained in terms of modality specific object representations then your experience does not represent the person you see as the source of the sounds: your perceptual experience does not represent a single object by means of a single representation of it as having visual features and as the source of the sounds; instead, it represents the object by means of one auditory representation which represents it as the source of the sound, and by another visual representation that represents the person, and each of these representations represents the (matching) spatio-temporal features of the object. These representations, and hence your perceptual experience, is veridical if the sounds come from the same place as the person you see, even if the person you see is not the source of the sounds.

Which account of the veridicality conditions is correct will have further consequences for our view of perception and the senses, some of which I'll mention later. But now, having sketched the two models, I want to turn to the question of how we determine which of them is correct. An obvious place to start is with empirical studies of inter-sensory interactions.

*

There are inter-sensory interactions that don't involve the kind of integration of information across object-representations described by the two models. Some of the ways in which the senses influence each other can be explained in terms of one sense modality

[19] There could be modality specific representations of objects and – in virtue of the operation of some supramodal perceptual or attentional mechanism – amodal representations of objects; or there could be modality specific *perceptual* representations of objects, but amodal representations of objects for guiding *action*. I am ignoring these to focus on the possibility and consequences of amodal perceptual representations of objects.

bringing about or causing a change in another sense modality, without appeal to representations of objects as such.

For example, there are processes that are responsible for calibrating spatial frames of reference across different sensory systems. In order to act on objects we can see, the bodily frame of reference that guides reaching must be aligned with the frame of reference relative to which the things we see are located. One kind of mechanism by which this calibration might occur uses optic flow – a pattern of change that can be specified at the level of the retina – to calibrate the direction of movement of the body within a visual frame of reference. This kind of calibration mechanism doesn't require recognising that something perceived with one sense (kinaesthesia) is the same as something perceived with another (vision), so doesn't require information about particular objects to be integrated. Instead, information about the alignment of the spatial frames of reference in different sensory systems is available, and can be used to produce a general calibration across senses, without reference to particular objects.[20]

In some cases an interaction that appears prima facie to involve object representations of particular things in fact doesn't. A single brief visual flash accompanied by two auditory beeps can result in the illusion of two flashes having occurred.[21] Conversely, a double flash accompanied by a single beep may be misperceived as a single flash.[22] In both cases, the auditory experience alters the visual experience of the flash to produce an illusion. This illusion occurs as a result of low-level connections between the auditory and visual cortex that enable activity in the auditory cortex to modulate activity in the visual cortex. In this case the modulation is temporal, and the auditory input changes the temporal properties of visual features. That the activity in the visual cortex occurs at the same time as activity in the auditory cortex indicates that it is likely to have been produced by the same environmental event. Given that likelihood, and the fact that the auditory resolution of time is more accurate than the visual, this kind of modulation may function to enhance the accuracy of the visual perception of brief environmental events.[23]

Although this example involves an interaction between the processes that ultimately produce perceptions of particular objects, it can be explained in terms of a mechanism that doesn't involve the integration or combination of information concerning particular objects. Activity at a low-level in one sensory system modulates the activity at a low-level in another sensory system. This results in changes in the properties of low-level feature detectors in a way that generally enhances perceptual performance.[24] But this interaction doesn't require

---

[20] Bruggeman, Zosh, and Warren, 2007.

[21] The illusion can be produced by briefly displaying a white disk against a black background, accompanied by a series of auditory beeps. Subjects who are asked to report how many flashes they see incorrectly report seeing a multiple flashes when a single flash is accompanied by more than one beep. See Shams, Kamitani, and Shimojo, 2000.

[22] Watkins et al. 2007.

[23] Watkins et al. 2006.

[24] Although it's not clear why these low-level connections produce the illusions, it may be that the visual flashes are close to the temporal threshold of what is visually perceivable. Modulation of low-

temporal information from vision and audition to be integrated or combined; and, since these early stages of perceptual processing don't involve representations of particular things, it would be implausible to suppose that it involves the integration of features associated with distinct object representations.

Although these kinds of interaction help co-ordinate and improve the performance of different sensory systems when they operate together, they do not involve either crossmodal or amodal perception. They can be explained in terms of the causal influence of modality specific sensory processes: what happens in one sensory system causes changes in another sensory system. Not all multi-sensory perception can be explained in this way.

*

One of the most familiar kinds of inter-sensory interaction is the ventriloquism effect. When a ventriloquist speaks without moving her lips, her voice seems to come from the mouth of the dummy whose moving mouth visually appears to be the source of the sounds. The ventriloquism effect produces an apparent change in the location of the source of the sound towards what visually appears to be the source of the sound. The effect occurs even for simple stimuli such as light flashes and tones: perceivers generally misjudge the location of a sound source if they hear the sound at the same time as they see a flash of light at a different (but nearby) location.

A similar phenomenon occurs in the temporal domain. In one experiment, subjects had to judge the order in which two small lights, arranged one above the other, were illuminated. Brief sounds were played from a loudspeaker behind the lights. On some trials there were no sounds or the sounds occurred simultaneously with the lights; on others one tone was played before the first light was illuminated and the second after the second light was illuminated. The sounds played before and after the lights led to an improvement in the subjects' performance. In another experiment, the first sound was played after the first light and before the second light, and the subjects' performance was worse than it was without the sounds. The sounds appear to have produced a temporal ventriloquism effect by 'pulling' the lights into temporal alignment with the sounds and so either increasing or reducing the apparent temporal separation between them, and therefore improving or reducing the subjects' ability to judge their temporal order.[25]

Something similar to the ventriloquism effect can occur for the auditory and visual perception of movement. This is demonstrated in the case of subjects who had to determine

---

level visual attention mechanisms may resolve what is, in effect, an ambiguous visual stimulus in a way that normally improves the reliability of visual perception, without there being integration of information. See Macaluso 2006.

[25] Morein-Zamir, Soto-Faraco, and Kingstone 2003. The temporal ventriloquism effect shows that there doesn't have to be precise temporal synchronisation across different senses for information to be integrated. Given data from a temporal order judgment task it is possible to determine how big the temporal interval between the stimuli can be for them to still be perceived as occurring at the same time (or the interval for which the subject is as likely to judge that the first stimulus a came before the second as they are that the second came before the first).

the apparent direction of the motion of a sound source, whilst ignoring the apparent motion of a light. The light could appear to move in a direction that was either congruent or incongruent with the apparent motion of the sound. The results "demonstrate a strong crossmodal interaction in the domain of motion perception… [they] suggest the obligatory perceptual integration of dynamic information across sensory modalities, often producing an illusory reversal of the true direction of the auditory apparent motion".[26]

Why do these different kinds of ventriloquism effect occur? In general, vision provides more accurate and more reliable spatial information than hearing, and hearing provides more accurate and more reliable temporal information than vision. When information about spatial and temporal features across different sense modalities conflicts, the perceptual system combines or integrates it in a way that favours the most accurate and reliable source of information. Greater weight is given to whichever source of information is most reliable in the circumstances. If visual information about spatial location is poor (as it might be in poor visibility) more weight is given to auditory spatial information; normally, however, (in good visibility) more weight is given to visual information. Integration involves the "near optimal combination of visual and auditory space cues, where each one is weighted by an inverse estimate of the noisiness, rather than one modality capturing the other".[27] The result is that the perception of spatial and temporal features of objects and events perceived with both vision and hearing is more accurate than it would have been if no integration occurred:[28] by integrating two or more sources of information, the variance inherent in each is reduced.[29] Multi-sensory perception is, therefore, more reliable and more accurate than perception with a single sense modality.

What does ventriloquism show about the nature of integration? Unlike the low-level interactions of the kind required to explain the flashing lights illusion, these effects can only be explained on the assumption that information concerning the same object, perceived with different senses, is integrated or combined. Why? Because the apparent location of an object – where the object perceptually appears to be – is the result of the optimal combination of information from different senses. The process of combining information must take as input the location of the object as represented in the auditory system and the location of the object as represented in the visual system, together with some estimation of the reliability of each, to

---

[26] Soto-Faraco et al. 2002, p.145. Does the perception of motion in these cases involve a temporal mismatch – with the time of a flash incongruent with the time of a sound – or a spatial mismatch – with the location of flash incongruent with the location of a sound? It could perhaps be either, or neither. The information integrated may be sense-specific information about direction of movement.
[27] Alais and Burr, 2004, p.260. This kind of optimal integration occurs across other senses too. Vision is more precise for discriminations along the horizontal, proprioception is a more reliable for discriminations in depth. There is evidence that the perceptual system takes this into account, giving extra weight to information from proprioception when the task requires depth discrimination, and extra weight to vision when it requires discrimination along the horizontal. It's not the case that vision always dominates touch; it does so only when it is the more reliable source of information. See Ernst and Bülthoff, 2004.
[28] See Alais and Burr, 2004, and Battaglia, Jacobs, and Aslin, 2003.
[29] Ernst and Bülthoff, 2004.

produce an (optimal) representation of the location of the object. The consequences of spatial ventriloquism cannot be explained simply in terms of one sensory system causally modulating the operation of the other: it involves the combination of the spatial information about an object from two different senses. The same argument applies in the temporal case, and in the case of motion.

Furthermore, in the temporal case it appears that there is a 'temporal window' within which audio-visual integration can occur. This window is asymmetrical and flexible:[30] its size changes according to the distance of a visually apparent sound source from the observer, and is wider for sounds from a more distant event. This suggests that the perceptual system is able to compensate for the fact that light from an object reaches the observer before the sound it makes, and that the gap between the two is greater for a sound from a more distant object. The properties of the temporal window can be explained in terms of the perceptual system compensating for this discrepancy in the perceived properties of particular objects; it can't be explained in terms of causal modulation.[31]

The very fact that inter-sensory integration involves the integration of a range of features associated with a distal or environmental object perceived with more than one sense might be taken to support the amodal model. Multi-sensory perception can be viewed as the parallel processing of information about features of objects detected by different sensory modalities. That means there is an inter-sensory binding problem analogous to the binding problem in visual perception. In visual perception, features in parallel processing streams that correspond to the same distal object must be grouped or 'bound' together into states that correspond to that object. In multi-sensory perception features in different sensory modalities that correspond to the same object must treated as belonging to the same distal object. Inter-sensory integration occurs when features in different sensory modalities that correspond to the same object are treated as features of a single object. So, it might be suggested, inter-sensory integration produces states that represent groups of features as belonging to particular distal objects; that is, inter-sensory integration produces amodal representations of objects, and so rules out the crossmodal model. O'Callaghan, in a recent discussion, suggests that inter-sensory integration "shows that there is a subpersonal grasp, at the level of sensory or perceptual processing, of sources of stimulation that must be understood in multi-modal or modality-independent terms. If you are willing to attribute content to subpersonal perceptual states, the corresponding states possess multi-modal content."[32] If that's right, then inter-sensory integration implies the existence of amodal representations.

But that conclusion is drawn too quickly. We can explain the different kinds of ventriloquism effect in a way that doesn't involve amodal object representations. For inter-

[30] Integration occurs when visual stimuli lead auditory stimuli by up to about 300ms, or lag by 80ms; subjects find it more difficult to detect asynchrony when the visual signal leads, than when the auditory signal leads.
[31] Though the process breaks down for distances of more than about ten metres (Sugita and Suzuki, 2003). For a survey of temporal ventriloquism, see Vatakis and Spence 2010.
[32] O'Callaghan 2007, 14.

sensory integration to occur, features of the same object perceived with different senses must be identified so that they can be integrated. The perceptual system makes use of a number of different cues in order to determine whether information across senses is likely to have come from the same object. Some of these cues are bottom-up and rely on correspondences between sensory features in different processing streams. For example, representations along early stages of visual processing may encode features such as changes in luminance and changes in motion; those along the early stages of auditory processing stages encode changes in intensity and changes in pitch and motion cues; these features correspond to the same properties of distal objects. If these features are correlated in time (and perhaps in space), they are likely to correspond to features of a single distal object. The perceptual system can exploit this, and treat features that are correlated in this way as corresponding to a single distal object.

Other cues are top-down, and draw on the subject's semantic or associational knowledge – on whether the subject takes, or is likely to take, what is perceived with two different senses to be a single distal object, or to be features of a single distal object. This is often labelled 'the unity assumption': "the assumption that a perceiver makes about whether he or she is observing a single multisensory event rather than multiple separate unisensory events."[33] Making this assumption needn't involve explicitly judging that what is perceived with one sense is the same as what is perceived with another.[34] It might result from a past association in experience of features associated with a single object, or from knowledge that certain features are likely to go together, in the way that the visual appearance of a steam kettle goes together with the whistling sound it makes. The perceptual system is able to exploit this knowledge, and treats features that are 'assumed' to belong together as corresponding to a single object. (Conversely, when features are 'assumed' not to belong together – when the unity assumption is false – the perceptual system treats them as belonging to distinct objects.)[35] In most circumstances, both top-down and bottom-up cues are likely to operate at the same time, with the result that information in different senses is integrated if and only if it is likely to have a single distal source.[36]

Integration does not only occur on a feature-by-feature basis. The fact that a feature in one sense is treated as corresponding to the same object as a feature in another sense means that information about that feature is integrated across the senses, but it also makes it likely that information about other features associated with that distal object are integrated across the senses. For example, if there are cues to indicate that spatial features in two senses are likely to be features of a single object, then information about those spatial features is integrated; but that makes it likely that information about the temporal features associated with the object is integrated too. So when one feature is treated as belonging to the same

[33] Vatakis and Spence 2007, 744.
[34] But what we judge or know does affect whether cross-modal integration occurs.
[35] See Vatakis and Spence 2007;Vatakis and Spence 2008; and Vatakis and Spence 2010.
[36] Vatakis and Spence 2007, 753.

single object across senses, then other features associated with that object are treated as belonging together across the senses.

This is nicely illustrated by the following demonstration of the ventriloquism effect. When a speech recording was played with a video recording of a talking head, inter-sensory effects were modulated by whether the gender of the voice matched that of the head in the video. "Participants found it significantly easier to discriminate the temporal order of the auditory and visual speech stimuli when they were mismatched than when they were matched [i.e. when the gender of the voice didn't match rather than matched that of the video]." These results don't just provide "psychophysical evidence that the 'unity assumption' can modulate crossmodal binding of multisensory information at a perceptual level,"[37] they show that when one – in this case, high-level – feature is treated as belonging to the same object across senses, other features associated with that object are treated as belonging to the same object and so are integrated.

In some cases, then, the fact that a feature represented by two distinct object representations is likely be a feature of the same distal object results in the integration of information about that feature across the representations; but it also leads to the integration of information about other features represented by the two object representations. The fact that one feature is integrated across senses as coming from the same distal object leads to the integration of other features of the same object. We cannot explain that in terms of the association of features across senses: the two object representations that represent those features must be associated. But that association of two object representations is consistent with the crossmodal model. An object representation in one sensory system can be associated with an object representation in another sensory system – on the basis of the kinds of cues that I have described – so that information about some of the features of the distal object that they represent can be integrated, without the two object representations being merged into a single object representation, and so without the formation of a single amodal object representation.

For example, in the ventriloquism effect we see the dummy's moving mouth and hear speech. We might suppose that the visual-object representation of the moving mouth and the auditory-object representation of the source of the speech are associated on the basis of shared temporal properties. As a consequence information about spatial features represented by the auditory object representation is integrated with information about the spatial features represented by the visual-object representation, so that we hear the speech to come from the place that we see mouth of the dummy to be. But this could happen without the spatial features losing their identity as features represented by two distinct object representations, and without the other features represented by the two object representations being merged into a single amodal object representation.

[37] Vatakis and Spence 2007, 752.

If that's right, then inter-sensory integration – at least of the kind involved in the ventriloquism effects – requires associating object representations of a single distal object across sensory modalities so that information about *some* features can be integrated, but it does not require that *all* the features associated with the two object representations are merged into a single object representation, and so does not require amodal object representations.

I argued that, whilst there's a sense in which the senses are modal, multi-sensory perception could not simply consist in the combined operation of each of the individual senses. The kind of inter-sensory integration required to explain the ventriloquism effects substantiates that conclusion: when the same thing is perceived with two or more senses the senses interact with each other in such a way that what is perceived with one sense can only be explained by appeal, in part, to what is perceived with the other senses. But this kind of integration does not undermine the idea that perceptual states are modality specific: that although the representational contents of the perceptual states of one modality are influenced by the contents of the perceptual states of another modality, our perceptual awareness of distal objects is explained in terms of modality specific object representations.

*

If the kind of inter-sensory integration involved in the various kinds of ventriloquism effects don't give us any reason to reject the crossmodal model, then what would? That is, what would show that inter-sensory interactions involved amodal object representations? According to the amodal model, when a single distal object is perceived with two senses and inter-sensory integration occurs, all the information about that distal object is combined into a single object representation – into a single amodal 'object file'. This amodal object file will contain information about all the features associated with the distal object, features that are perceived with more than one sense modality. So it will contain information about both auditory and visual features. On the other hand, according to the crossmodal model there are distinct sense modality specific object representations – sense-specific 'object files' – associated with each sense modality. Although information concerning some features – those perceived with both senses – is integrated across these representations, each representation contains information about features specific to that sensory modality. So the visual-object representation will represent visual features of the distal object that the auditory-object representation doesn't, and vice versa.

That suggests that we can test the claim that there are amodal object representations by testing whether information about sense specific features from more than one sense is contained within a single object file. I described how the existence of an object specific preview benefit provides evidence that visual information about different features is grouped into an object file. If information about features specific to more than one sense modality is contained within a single object file, then we would expect there to be crossmodal object specific preview benefits.

I have not been able to find many attempts to test this suggestion, but in one experiment subjects saw two object targets that briefly displayed pictures (of a dog, whistle, train,

hammer, piano, and phone); the objects then moved to a new position and a sound (of a dog bark, a whistle blow, a train horn, a hammer blow, a piano note, and a phone ring) was played from a position that corresponded to the position of one of the objects. In some trials the sound played was one that matched the picture previously displayed on the object, and in others the sound played was one that matched the picture previously displayed on the other object. Subjects had to report whether what they heard corresponded to what they had previously seen. An object specific previous benefit was found.[38]

This result is just what we would expect if inter-sensory integration involves amodal object files. The subject produces a response on the basis of auditorily perceiving an object (hearing the sound); the fact that there is a preview benefit suggests that the auditory object file contains information contained in the visual object file associated with the object they saw moving; so there is a single file that contains both auditory and visual information as I suggested there would be if the amodal model is correct. Of course, this is only a single experiment, and it doesn't show that *all* the information associated with the object is contained within a single file. But I think it is difficult to explain these results on the assumption that there are sense-modality specific object representations.

A different kind of evidence comes from the fact that auditory experience can tell us about changes to, or events occurring in, objects that are not otherwise visible. To take a simple example, if a box on your desk is making a ticking sound then you can perceive something is happening to it even though there is nothing visibly happening to it. The two models of inter-sensory integration give different accounts of what you perceive in this kind of case.

If there are amodal objects representations, then auditory information about what is happening to a distal object will be combined with visual information into a single amodal object representation. Such a representation will not just contain information that an event occurred, but that an event occurred to a visible distal object. So you perceive something happening to the thing you can see. If there are distinct, sense-modality specific object representations, then auditory information about what is happening to a distal object will be contained in an object representation that is distinct from the visual-object representation of the same object. There will be an auditory representation of an event, and a distinct visual representation of an object. These representations might represent some of the same features of the distal object – the same spatial features for example – but the auditory features and visual features are not represented as features of a single object: there is not a single amodal representation of an object as having both auditory and visual features. So although you hear something happening at the same place as an object that you can see, you do not hear something happening to the object that you can see.

The occurrence of an illusion in which auditory information disambiguates a visually ambiguous display provides empirical evidence in favour of the amodal account of this kind of case.[39] The illusion involves two objects each of which start out at the top corner of a

[38] Jordan, Clark, and Mitroff 2010.
[39] Sekuler, Sekuler, and Lau 1997.

screen and then move diagonally to the bottom corner, crossing in the centre. The display is ambiguous. It can be perceived either as the objects colliding with each other in the centre of the screen and rebounding (as if the objects had 'bounced' off each other), or as each object following a straight path to the opposite bottom corner and so 'streaming' past the other in the centre of the screen. If an auditory event – a tone or tap – coincides with the moment they meet, the objects are seen as bouncing. The auditory event disambiguates the display. The disambiguating effect of the sound (the auditory event) is reduced if it is flanked by other sounds. That seems to be because when it is flanked by other sounds the sound is perceived as part of a distinct auditory object (the stronger the grouping of the sound with the flanking sounds, the less likely it is to have a disambiguating effect).[40]

The experiences of the display differ according to whether or not an event – a collision of the objects – is perceived to occur. A collision is perceived to occur when the meeting of the objects in the centre of the screen is perceived as the source of the sound – as a collision that produces the sound. Since the sound's source has the same spatio-temporal properties in all cases, but only disambiguates the display in some, it is not sufficient for the meeting of the objects to be perceived as the source of the sound that it is perceived to have the same spatio-temporal properties as the source of the sound. It is not possible to explain the disambiguation as a spatio-temporal effect.

When the source of the sound is perceived to be something that produces a sequence of sounds (that is, when the sound is perceived as part of a distinct auditory object) – and so to be unrelated to the meeting of the objects – no collision is perceived to occur. It seems, then, that the best explanation of the perception of a collision is that the source of the sound is represented as an event involving the meeting of the two objects; that is, an event is represented as both the meeting of the two objects and as the source of the sound. Such a representation would not just contain information that an event occurred, but that an event occurred to a visible object. It would be an amodal representation. Since the best explanation of this illusion is in terms of the amodal representation of an event, the occurrence of the illusion supports the amodal model of multimodal perception.

*

I have described two kinds of evidence that lend support the suggestion the multi-sensory perception involves the amodal representations of objects, but clearly more is needed. Representing something as an object involves representing it as cohesive, bounded, and spatio-temporally continuous, and that involves being able to keep track of it over time and through changes. If objects are represented amodally then capacities to track objects over time and through changes will be amodal capacities. For example, an object might be tracked initially by hearing it, and then by later by seeing it. In vision, the capacity to keep track of an object visually consists, partly, in perceptual anticipation: what is perceived at one time has consequences for what will be perceived at a later time. Does perceiving an object

[40] Watanabe and Shimojo 2001.

with one sensory modaltiy generate perceptual anticipations in other senses? If the capacity is amodal, we might expect hearing an object to generate visual expectations. The literature on child development contains many elegant experiments that probe the properties of infants' visual representations of object, including their ability to visually track objects over time. It would be interesting to discover whether their ability to track objects is amodal: for example, if an infant *hears* something behind a barrier, would they be surprised, when the barrier is removed, to *see* nothing there?[41]

*

I began by arguing that multi-sensory perception cannot simply be the combined operation of each of the individual senses, but that some kind of integration of information across the senses is required. I sketched two different models of what that integration of information might involve. Many examples of multi-sensory perception are consistent with both models, but I have described some evidence that supports the suggestion that some multi-sensory perception involves the amodal representation of distal objects. Much of our theorising about perception takes place within a modality specific framework. We focus on vision, or on audition, and consider it in isolation from the other senses. If perception is amodal, then such theorising is likely to leave out something fundamental about the nature of perception – that it involves capacities that are shared across different senses. For example, considered in isolation from vision, auditory perception can appear limited in what it can tell us about the world, perhaps limited to telling us about sounds and their features. But considered in conjunction with vision, as a form of perception that draws on the same capacities to perceive objects as vision does, auditory perception appears a far less limited – a form of perception that enables us to perceive the same material objects that we see. Any account of auditory perception must explain how vision and the other senses – and capacities shared across the senses – contribute to what is auditorily perceived; and any account that doesn't do so will leave out something central to auditory perception. If perception draws on amodal representational capacities, then that argument applies generally: just as audition cannot be understood in isolation from vision, so vision cannot be understood in isolation from audition.[42]

Alais, David, and David Burr, "The ventriloquist effect results from near-optimal bimodal integration," *Current Biology: CB* 14, no. 3 (February 3, 2004): 257-262.

Battaglia, Peter W, Robert A Jacobs, and Richard N Aslin, "Bayesian integration of visual

[41] Richardson and Kirkham (2004) used a multimodal object tracking experiment that may lend support to the suggestion that that they would, but is perhaps better explained as a semantic rather than a perceptual effect.

[42] I would like to thank the editors for many helpful suggestions.

and auditory signals for spatial localization," *Journal of the Optical Society of America. A, Optics, Image Science, and Vision* 20, no. 7 (July 2003): 1391-1397.

Bregman, Albert. *Auditory Scene Analysis*. Cambridge, Ma.: MIT Press, 1994.

Bruggeman, Hugo, Wendy Zosh, and William H. Warren, "Optic Flow Drives Human Visuo-Locomotor Adaptation," *Current Biology* 17, no. 23 (2007): 2035-2040

Burge, Tyler. *Origins of Objectivity*. Oxford: Oxford University Press, 2010.

Casati, Roberto, and Jerome Dokic. "Sounds." Stanford encyclopedia of philosophy (Fall Edition), ed. Edward N. Zalta. (2005).
URL=<http://plato.stanford.edu/archives/fall2005/entries/sounds/>.

Ernst, Marc O and Heinrich H Bülthoff, "Merging the senses into a robust percept," *Trends in Cognitive Sciences* 8, no. 4 (April 2004): 162-169.

Fodor, Jerry A. *The Modularity of Mind*. Cambridge, MA.: MIT Press (1983).

Jordan, Kerry E., Kait Clark, and Stephen R. Mitroff, "See an object, hear an object file: Object correspondence transcends sensory modality," *Visual Cognition* 18, no. 4 (2010): 492.

Macaluso, Emiliano. "Multisensory Processing in Sensory-Specific Cortical Areas," *The Neuroscientist* 12, no. 4 (2006): 327 −338.)

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

Matthen, Mohan. "On the Diversity of Auditory Objects." *Review of Philosophy and Psychology* 1 (2010): 63-89.

Morein-Zamir, Sharon, Salvador Soto-Faraco, and Alan Kingstone, "Auditory capture of vision: examining temporal ventriloquism," *Cognitive Brain Research* 17, no. 1 (June 2003): 154-163.

Nudds, Matthew. 'Sounds and Space.' In Nudds and O'Callaghan (2010): pp-pp

Nudds, Matthew. "What are auditory objects?" *Review of Philosophy and Psychology* 1 (2010): 105-122.

Nudds, Matthew. "The Senses as Psychological Kinds" in Fiona Macpherson ed. *The Senses: Classical and Contemporary Readings*. Oxford: Oxford University Press (2011): 311-40.

Nudds, Matthew. "Auditory Perception." In Mohan Mathen, ed. *Oxford Handbook of Perception.* Oxford: Oxford University Press (forthcoming).

Nudds, Matthew, and Casey O'Callaghan. *Sounds and Perception: New Philosophical Essays*. Oxford: Oxford University Press (2010).

O'Callaghan, Casey. "Perception and Multimodality." MS.

O'Callaghan, Casey. *Sounds: A Philosophical Theory*. Oxford: Oxford University Press, 2007
Palmer, S. *Vision Science: From Photons to Phenomenology*. Cambridge, Ma.: MIT Press, 1999.

Pasnau, Robert. "What is sound?" *Philosophical Quarterly* 50 (1999): 309-24.
Richardson, Daniel C., and Natasha Z Kirkham, "Multimodal events and moving locations: eye movements of adults and 6-month-olds reveal dynamic spatial indexing," *Journal of Experimental Psychology, General* 133, no. 1 (March 2004): 46-62.
Sekuler, R., AB Sekuler, and R Lau, "Sound alters visual motion perception," *Nature* 385, no. 6614 (January 23, 1997): 308

Shallice, Tim. *From Neuropsychology to Mental Structure*. Cambridge: Cambridge University Press (1988).

Shams, Ladan, Yukiyasu Kamitani, and Shinsuke Shimojo, "Illusions: What you see is what you hear," *Nature* 408, no. 6814 (December 14, 2000): 788.

Soto-Faraco, Salvador, et al., "The ventriloquist in motion: Illusory capture of dynamic information across sensory modalities," *Cognitive Brain Research* 14, no. 1 (June 2002): 139-146.

Sugita, Yoichi and Yôiti Suzuki, "Audiovisual perception: Implicit estimation of sound-arrival time," *Nature* 421, no. 6926 (February 27, 2003): 911.

Treisman, A. "Feature binding, attention and object perception," *Philosophical Transactions of the Royal Society B: Biological Sciences* 353, no. 1373 (August 29, 1998): 1295-1306.

Vatakis, Argiro, and Charles Spence, "Crossmodal binding: evaluating the 'unity assumption' using audiovisual speech stimuli," *Perception & Psychophysics* 69, no. 5 (July 2007): 744-756.

Vatakis, Argiro, and Charles Spence, "Evaluating the influence of the 'unity assumption' on the temporal perception of realistic audiovisual stimuli," *Acta Psychologica* 127, no. 1 (January 2008): 12-23.

Vatakis, Argiro, and Charles Spence, "Audiovisual Temporal Integration for Complex Speech, Object-Action, Animal Call, and Musical Stimuli," in *Multisensory Object Perception in the Primate Brain*, ed. Jochen Kaiser and Marcus Johannes Naumer (Springer New York, 2010), 95-121.

Watanabe, K., and S Shimojo, "When sound affects vision: effects of auditory grouping on visual motion perception," *Psychological Science: A Journal of the American Psychological Society / APS* 12, no. 2 (March 2001): 109-116.

Watkins, S. et al., "Sound alters activity in human V1 in association with illusory visual perception," *NeuroImage* 31, no. 3 (July 1, 2006): 1247-1256.

Watkins, S. et al., "Activity in human V1 follows multisensory perception," *NeuroImage* 37, no. 2 (August 15, 2007): 572-578.

Zmigrod, S. and B. Hommel, "Auditory event files: Integrating auditory perception and action planning," *Attention, Perception & Psychophysics* 71, no. 2 (February 2009): 352-362.